Choice Evaluation and Context Effects

by

Takao Noguchi

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Psychology

May 2014

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank Adam Sanborn and Neil Stewart for their supportive supervision of this thesis. Their advice and comments were invaluable. I am also very grateful to Thomas T. Hills, with whom I have collaborated on some of the projects presented here.

I was financially supported by the Department of Psychology.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data collection and data analysis) was carried out by the author except in the cases outlined below: Chapter 2 was written in collaboration with Adam Sanborn and Neil Stewart; Chapters 3 and 4 were written in collaboration with Thomas T. Hills; and Chapters 5 and 6 were written in collaboration with Neil Stewart.

List of publications including submitted papers:

Chapter 2

Noguchi, T., Sanborn, A., & Stewart, N. (2013). Non-parametric estimation of the individual's utility map. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Chapter 3

Noguchi, T., & Hills, T. T. (submitted). Set-size induced risk-amplification: Experience-based decisions in large set-sizes favour riskier alternatives.

Noguchi, T., & T. T. Hills. (2014). Context effects and risk amplification: Why more is risky. In P. Bello, M. Guarini, M. McShane, and B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Chapter 4

Noguchi, T., & Hills, T. T. (submitted). Thin-search and choice deferral: Encouraging myopic choices with more information.

<u>*Chapter 5*</u>

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition, 132*, 44–56.

Abstract

Behavioral research has long documented that the choices an individual makes do not always follow the maximization of expected values. To describe the utility an individual maximizes through his or her choices, one class of models — static models — has been previously developed. These models are reviewed in Chapter 1. To assess the static models, a non-parametric method to reveal the utility of alternatives is developed in Chapter 2. The results show that the utility predicted from the static models deviates from the estimated utility.

Utility, however, is relatively unstable across contexts determined by information presentation formats, choice set-sizes, the structures of alternatives, and the relationships between alternatives. This instability is a topic for Chapters 3, 4, and 5. Following Chapter 3, which examines effects of information presentation formats and choice set-sizes on risk-taking, Chapter 4 further investigates how the contexts impact on choice evaluation. Then, Chapter 5 examines process of choice evaluation by analyzing eye-movements during choices. The results from these three chapters indicate that choices are systematically altered with contexts, supporting instability of utility.

The instability of utility conflicts with the principle of utility maximization, and Chapters 5 and 6 consider another class of models — dynamic models — which can accommodate utility instability. A dynamic model assumes that an individual iteratively and stochastically develops preferences for each alternative, until preference for one alternative reaches a choice criterion. The exact processes of preference development is investigated in Chapter 5, which suggests that a dynamic model should be based on single-attribute pair-wise comparisons. Following this suggestion, a new model — multi-alternative decision by sampling — is proposed in Chapter 6.

Chapter 7 discusses overall implications of the results for the principle of utility maximization and model evaluation. I conclude that models should be assessed not only on their ability to predict choices but also on their ability to predict concurrent process measures, including eye-movements.

Chapter 1

Introduction

Choice making lies at the heart of human behavior. When an individual walks across a busy street, the individual is assessing a risk of being run over a car and choosing when to cross the street. When an individual looks through a menu at a restaurant, the individual is evaluating each dish on the menu and choosing which dish he or she prefers the most. Some of choices, such as choosing who to marry with, have life-changing consequences, while others may have more trivial consequences. Studies on how an individual makes those choices has been of great interest for psychologists, economists and researchers in many other field. Psychology of choice has been studied since the 18th century and has attracted increasing attention especially over the last three decades.

This thesis investigates how an individual evaluates an alternative and how he or she makes a choice, and this investigation extensively uses mathematical models to understand choices. There exist numerous mathematical models to explain a number of phenomena observed with various experimental paradigms, ranging from perceptual choice (e.g., choosing a brighter patch) to consumer choice (e.g., choosing a car to purchase). These mathematical models have been practically and theoretically proved useful, as they can be readily tested with various alternatives and compared against actual choice behavior.

Many of the existing models can be classified into the class of static models or the class of dynamic models. Static models are often developed in behavioral economics to explain choice between probabilistic pay-offs. In contrast, dynamic models typically stem from psychophysics to explain speed and accuracy of perceptual judgments and have been extended to explain choice between consumer products. Although other classes of models exist in the literature (e.g., heuristic models), this introductory chapter focuses on the classes of static and dynamic models, especially the models I use throughout the rest of thesis. Thus, this chapter does not aim to provide comprehensive review of the field, but rather, it aims to introduce the models relevant to the chapters to follow.

1.1 Static Models

Choice models, especially those in economics, often assume utility maximization. Here, an individual is assumed to choose an alternative which maximizes utility. Since studies in 18th century, it has been empirically shown that an individual's choice is not well explained by maximization of expected pay-off. To understand an individual's choice, various models have been proposed to describe utility of an alternative. I start the review with one of the earliest models: expected utility theory.

1.1.1 Expected utility theory

Expected utility theory was first proposed by Bernoulli (1738/1954) to explain behavioral phenomena associated with a price an individual should pay to play a gamble. The individual can be assumed to pay any price up to the expected pay-off from the gamble, but violation of this assumption is observed with St. Petersburg game. In this game, a coin is flipped repeatedly until a head is produced. If an individual enters this game, the individual earns pay-off of £2^n, where n is the number of coin flips before landing a first head. An expected pay-off from this game is calculated as follows:

$$\sum_{n=1}^{\infty} \frac{1}{2^n} 2^n = \sum_{n=1}^{\infty} 1 = \infty.$$

If an individual is willing to pay any price up to the expected pay-off, this individual should be willing to pay any price at all: for example, hundreds of pounds.

An individual is, however, typically willing to pay only a few pounds, violating the assumption that his or her choice is explained well with maximization of the expected pay-off. This violation led Bernoulli (1738/1954) to formulate expected utility theory. According to this theory, an individual maximizes expected utility — subjective value of pay-offs —, rather than the actual expected pay-offs. Thus, Bernoulli introduced an utility function, which converts monetary amount into subjective value. In order to explain behavior in the St. Petersburg game, this utility function has to be concave. The utility function suggested by Bernoulli (1738/1954) himself is the logarithmic function, but for the sake of brevity and consistency with

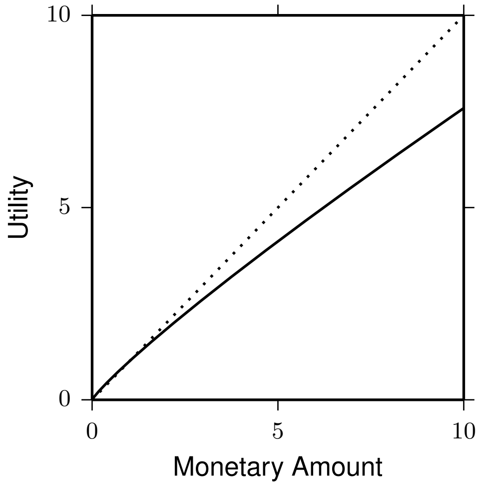


Figure 1.1: Example utility function: $u(x) = x^{0.88}$. Dotted line represents an identity function. Parameter value, 0.88, is taken from Tversky and Kahneman (1992).

the other models, I assume a power function, $u(x) = x^{\alpha}$, where $0 < \alpha < 1$. As an illustration, an utility function with $\alpha = 0.88$ is displayed in Figure 1.1. Then, expected utility in the St. Petersburg game is calculated as below:

$$\sum_{n=1}^{\infty} \frac{1}{2^n} u(2)^n = \sum_{n=1}^{\infty} \frac{1}{2^n} (2^{\alpha})^n = \sum_{n=1}^{\infty} \frac{1}{2^n} 2^{\alpha n} = \sum_{n=1}^{\infty} 2^{\alpha n - n}$$
$$< \sum_{n=1}^{\infty} 2^{n-n} = \sum_{n=1}^{\infty} 1 = \infty. \quad (\because \alpha < 1)$$

With a concave utility function, an expected utility in the St. Petersburg game is less than infinity, which predicts that an individual should be willing to pay only up to a finite amount of price to enter this game. Please note however, that utility function is not a general solution, since other pay-off structures could still lead to infinite value in a variant of St. Petersburg game.

1.1.2 Prospect theory

While expected utility theory successfully describes an individual's behavior in the St. Petersburg game, empirical studies from 1950s reported a variety of choice behavior inconsistent with expected utility theory. These violation of expected utility theory has been throughly reviewed by Birnbaum (2008), Loomes (2010), and Schoemaker (1982). Here, I will only review one of the most well-known violations:

Allais's paradox (Allais, 1953).

This paradox is empirically tested by Kahneman and Tversky (1979), where an individual is asked to make a choice between the following two alternatives.

A:	2,500 with probability	.33,	B:	2,400 with certainty.
	2,400 with probability	.66,		
	0 with probability	.01;		

Kahneman and Tversky (1979) report that 82% of individuals chose Alternative B over A. Kahneman and Tversky (1979) also tested the following two alternatives:

C:	2,500 with probability	.33,	D:	2,400 with probability	.34,
	0 with probability	.67;		0 with probability	.66.

These alternatives are obtained by eliminating a .66 probability of 2,400 pay-off from Alternatives A and B, and hence, ordering of expected utility should not be affected: if Alternative B is chosen over A, Alternative D should be chosen over C. However, 82% of individuals chose Alternative C over D. This pattern of behavior indicates violation of expected utility theory.

According to expected utility theory, the choice of Alternative B over A implies the following:

$$.33\, u(2,500) + .66\, u(2,400) < u(2,400),$$

which is equivalent to

$$.33\, u(2,500) < .34\, u(2,400).$$

The left term of this inequality corresponds to Alternatives C, and the right term corresponds to D. Thus, expected utility theory predicts that if Alternative B is chosen over A, Alternative D should be chosen over C. This prediction is violated in the observation by Kahneman and Tversky (1979), where Alternative C is more frequently chosen than D.

Prospect theory resolves this paradox and several other violations of expected utility theory with editing rules and a weighting function. The editing rules pre-process probabilities and pay-offs of an alternative before applying the utility and the weighting functions. To resolve Allais's paradox however, only the weighting function is required. With the weighting function, w, choice of Alternative B over A implies:

$$w(.33)\, u(2,500) + w(.66)\, u(2,400) < u(2,400),$$

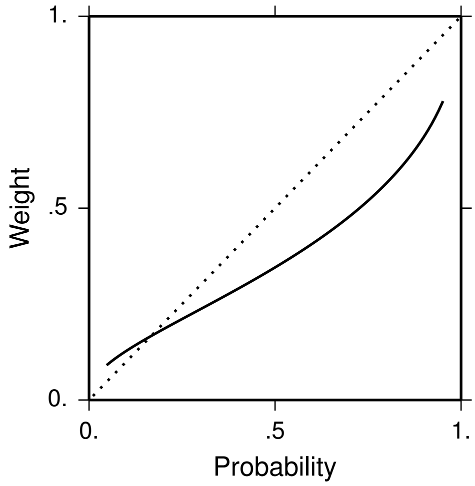


Figure 1.2: Example weighting function for prospect theory: $w(p) = \exp(-1.30\ (-\log(p))^{0.55})$. Dotted line represents an identity function. Functional form is taken from Prelec (1998), and parameter values are arbitrarily specified to reproduce the weighting function illustrated in Kahneman and Tversky (1979).

which is equivalent to

$$w(.33)\ u(2,500) < (1 - w(.66))\ u(2,400).$$

Also, choice of Alternative C over D implies

$$w(.34)\ u(2,400) < w(.33)\ u(2,500).$$

Thus if an individual chooses Alternative B over A and Alternative C over D, these choices implies

$$w(.34)\ u(2,400) < (1 - w(.66))\ u(2,400),$$

or

$$w(.34) < 1 - w(.66).$$

More generally, prospect theory is able to resolve Allais's paradox, as long as the following condition holds:

$$w(1 - p) < 1 - w(p). \tag{1.1}$$

This condition is satisfied by having a non-linear weighting function. Kahneman and Tversky (1979) do not provide a functional form for a weighting function, but Prelec (1998) examined several forms of a weighting function. An example is illustrated in Figure 1.2. Mathematical proof that this particular weighting function satisfies Equation 1.1 is provided by Prelec (1998).

Thus with the editing rules and a non-linear weighting function, prospect theory is able to explain Allais's paradox and various other behavior (see Kahneman & Tversky, 1979, for details), which expected utility theory is not able to explain. Also as prospect theory has an utility function u as expected utility theory does, prospect theory predicts that subjective value of the St. Petersburg game is finite.

1.1.3 Cumulative prospect theory

One major limitation of prospect theory, however, is that it potentially permits violation of stochastic dominance. Suppose an individual is making a choice between Alternatives E and F:

E:	x_1 with probability	p_1,	F:	x_1 with probability	p_1',
	x_2 with probability	p_2,		x_2 with probability	p_2',
	0 with probability	$1 - p_1 - p_2$;		0 with probability	$1 - p_1' - p_2'$.

Further assume that

$$\begin{aligned} 0 < x_1 < x_2, \\ p_2' < p_2, \quad \text{and} \\ p_1 + p_2 = p_1' + p_2' < 1. \end{aligned} \tag{1.2}$$

Then, Alternative E dominates F: Alternative E has a higher probability of obtaining a larger pay-off, x_2, than Alternative F, and also E has the same probability of obtaining the worse pay-off, 0, as F. Thus, Alternative E should be chosen over F, which implies

$$w(p_1')\, u(x_1) + w(p_2')\, u(x_2) < w(p_1)\, u(x_1) + w(p_2)\, u(x_2),$$

or

$$\frac{u(x_1)}{u(x_2)} < \frac{w(p_2) - w(p_2')}{w(p_1') - w(p_1)}. \tag{1.3}$$

Also, suppose the same individual is faced with a choice between Alternatives E' and F':

E':	x_1 with probability p_2',		F':	x_1 with probability p_2,	
	x_2 with probability p_1',			x_2 with probability p_1,	
	0 with probability $1 - p_2' - p_1'$;			0 with probability $1 - p_2 - p_1$.	

Alternative E' dominates F', because Equation 1.2 indicates $p_1 < p_1'$: Alternative E' has a higher probability of obtaining a larger pay-off than Alternative F'. Also, Alternative E' has the same probability of obtaining the worst pay-off as Alternative F'. Then, choice of Alternative E' implies the following:

$$w(p_2)\ u(x_1) + w(p_1)\ u(x_2) < w(p_2')\ u(x_1) + w(p_1')\ u(x_2),$$

or

$$\frac{u(x_1)}{u(x_2)} < \frac{w(p_1') - w(p_1)}{w(p_2) - w(p_2')}. \tag{1.4}$$

As x_1 becomes similar to x_2, the left term in Equations 1.3 and 1.4 approaches to 1. Thus, satisfaction of Equations 1.3 and 1.4 implies:

$$\begin{aligned}
& 1 \le \frac{w(p_2) - w(p_2')}{w(p_1') - w(p_1)} \text{ and } 1 \le \frac{w(p_1') - w(p_1)}{w(p_2) - w(p_2')} \\
\Leftrightarrow\quad & 1 \le \frac{w(p_2) - w(p_2')}{w(p_1') - w(p_1)} \text{ and } 1 \ge \frac{w(p_2) - w(p_2')}{w(p_1') - w(p_1)} \\
\Leftrightarrow\quad & 1 = \frac{w(p_2) - w(p_2')}{w(p_1') - w(p_1)} \\
\Leftrightarrow\quad & w(p_1) + w(p_2) = w(p_1') + w(p_2')
\end{aligned}$$

However, $w(p_1) + w(p_2)$ cannot be the same as $w(p_1') + w(p_2')$, as w is a non-linear function and $p_1 + p_2 = p_1' + p_2'$ is assumed (Equation 1.2). Thus when x_1 is very close to x_2, prospect theory eventually violates Equation 1.3 or 1.4, predicting a choice of the dominated alternative.

This limitation is overcome with cumulative prospect theory (Tversky & Kahneman, 1992), which employs a cumulative weighting function. Suppose an alternative is associated with a p_i probability of x_i pay-off, where $i = 1, 2, \ldots, n$. When $x_i < x_j$ for all i and j which satisfy $1 \le i < j \le n$, then a weight is computed as follows:

$$w(p_i) \quad = \quad \pi(\sum_{j=i}^{n} p_j) - \pi(\sum_{j=i+1}^{n} p_j), \text{ if } 1 \le i < n,$$

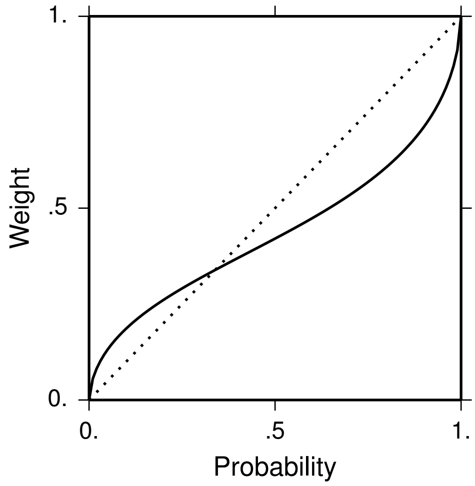


Figure 1.3: Example weighting function for cumulative prospect theory: $\pi(p) = \frac{p^{0.61}}{(p^{0.61}+(1-p)^{0.61})^{\frac{1}{0.61}}}$. Dotted line represents an identity function. Parameter value, 0.61, is taken from Tversky and Kahneman (1992).

and

$$w(p_n) \quad = \quad \pi(p_n).$$

With this cumulative weighting function, the right term in Equation 1.3 becomes the following:

$$\begin{aligned} \frac{w(p_2) - w(p_2')}{w(p_1') - w(p_1)} &= \frac{\pi(p_2) - \pi(p_2')}{\pi(p_2' + p_1') - \pi(p_2') - \pi(p_2 + p_1) + \pi(p_2)} \\ &= \frac{\pi(p_2) - \pi(p_2')}{-\pi(p_2') + \pi(p_2)} \quad (\because p_2 + p_1 = p_2' + p_1') \\ &= 1 \\ &> \frac{u(x_1)}{u(x_2)} \quad (\because x_2 > x_1). \end{aligned}$$

Thus, cumulative prospect theory satisfies both Equations 1.3 and 1.4, and hence stochastic dominance, independent of exact expression of π.

To explain choice better, however, Tversky and Kahneman (1992) proposes the following functional expression for π:

$$\pi(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}.$$

An example function with $\gamma = 0.61$ is displayed in Figure 1.3. This S-shape signifies typical risk-seeking for a small probability. An individual tends to overweight small probabilities and underweight high probabilities, so that this individual often chooses an alternative with a small probability of large pay-off over an alternative with a high probability of small pay-off. Also when this weighting function is linear, cumulative prospect theory becomes identical to expected utility theory. Thus, cumulative prospect theory can be seen as a general case of expected utility theory. Performance of cumulative prospect theory is compared against performance of expected utility theory in Chapter 2, and also an extension of cumulative prospect theory is used in simulations in Chapter 3.

1.1.4 Transfer of attention exchange model

More recent empirical studies, however, report notable violation of cumulative prospect theory (see Birnbaum, 2008, for review). Here, I review only one of such violations: violation of coalescing. In one of the experiments reported by Birnbaum (2008), an individual made a choice between the following two alternatives:

G:	100 with probability	.85,	H:	100 with probability	.95,
	50 with probability	.15;		7 with probability	.05.

The majority of individuals choose Alternative G over H. However, when confronted with the following two alternatives, the majority of individuals choose Alternative H' over G'.

G':	100 with probability	.85,	H':	100 with probability	.85,
	50 with probability	.10,		100 with probability	.10,
	50 with probability	.05;		7 with probability	.05.

Alternative G is a coalesced form of Alternative G'. The pay-off of 50 with a probability of .15 in Alternative G is split into the two possible pay-off of 50 in Alternative G'. Similarly, the pay-off of 100 with a probability of .95 is split into the two possible pay-offs in Alternative H'. In cumulative prospect theory, split pay-offs are automatically coalesced because of cumulativity of a weighting function, and thus, cumulative prospect theory predicts that if an individual chooses Alternative G over H, this individual should choose G' over H'. The violation is also reported by Birnbaum (1999).

To explain violation of coalescing, Birnbaum and Chavez (1997) propose transfer of attention exchange model. Suppose an alternative is associated with a p_i probability of x_i pay-off, where $i = 1, 2, \ldots, n$. When $x_i < x_j$ for all i and j which satisfy $1 \le i < j \le n$, then according to transfer of attention exchange model, subjective value of this alternative is computed as follows:

$$\frac{\sum_{i=1}^{n} t(p_i)\ u(x_i) + \sum_{i=1}^{n} \sum_{k=i}^{n} (u(x_i) - u(x_k))\ \ \omega(p_i, p_k, n)}{\sum_{i=1}^{n} t(p_i),}$$

where t is a weighting function, $t(p) = p^\gamma$, and

$$\omega(p_i, p_k, n) = \begin{cases} \dfrac{\delta\ t(p_k)}{n+1} & \delta > 0, \\ \\ \dfrac{\delta\ t(p_i)}{n+1} & \delta \leq 0. \end{cases}$$

The weight transfer function, ω, determines how much attention transfers from a better possible pay-off to a worse pay-off, where attention transfer is specified by parameter δ.

This transfer of attention exchange model does not coalesce pay-offs, and as a result, the model predicts different subjective values for Alternatives G than G'. As Alternative G has only one worse pay-off of 50 and Alternative G' has two, less attention transfers to 50 pay-off for Alternative G than G'. Thus, the coalescing makes Alternative G better than G'. Similarly, coalescing makes Alternative H worse than H', because more attention transfers from a better pay-off of 100 for Alternative H than for H'.

The transfer of attention exchange model explains Allais's paradox with transfer of attention. When evaluating the following alternatives,

A:	2,500 with probability	.33,	B:	2,400 with certainty,
	2,400 with probability	.66,		
	0 with probability	.01;		

attention transfers from pay-offs of 2,500 and 2,400 to pay-off of 0 in Alternative A. As a result, subjective value for Alternative A is less than subjective value for Alternative B. When 2,400 with probability .33 is subtracted from each Alternative, however, attention transfers to pay-off of 0 in both alternatives:

C:	2,500 with probability	.33,	D:	2,400 with probability	.34,
	0 with probability	.67;		0 with probability	.66.

As a result, the difference in the pay-offs of 2,500 and 2,400 carries over to the difference in subjective values of alternatives: Alternative C is predicted to be chosen by the transfer of attention exchange model.

In addition, the transfer of attention provides a different explanation for risk aversion than the other models reviewed above. Risk aversion refers to the empirical finding that an individual prefers a pay-off with probability 1.0 over a probabilistic pay-off with the same or even higher expected pay-off. An example is the following two alternatives:

K:	2,500 with probability	1.00;	L:	5,000 with probability	.50,
				0 with probability	.50.

Majority of individuals prefers Alternative K over L. This risk aversion is explained with a concave utility function in expected utility theory, which predicts that $u(2,500) \times 1.0 > u(5,000) \times 0.5$. Similar explanation is provided by the concave utility function together with the weighting function in prospect theory and cumulative prospect theory.

The transfer of attention exchange model, however, explains risk aversion with attention transfer: in Alternative L, attention transfers from 5,000 pay-off to 0 pay-off, resulting in less weight on 5,000 than .50. Thus with the identity utility function, the transfer of attention exchange model predicts that the subjective value of Alternative L is less than $5,000 \times 0.5 = 2,500$, which is the subjective value of Alternative K. The same mechanism predicts finite subjective value for the St. Petersburg game. As the transfer of attention exchange model can explain more choice phenomena, this model can be more flexible than other models.

Performance of the transfer of attention exchange model is assessed and compared against performance of expected utility theory and cumulative prospect theory in Chapter 2.

1.2 Dynamic Models

In contrast to static models, which have been developed in a domain of economic choice, dynamic models were initially developed with perceptual identification tasks. An example perceptual task presents two sequences of alphabetical characters and ask an individual to respond “yes” if the two sequences are both words or both non-words and “no” otherwise (Meyer & Irwin, 1981). To explain speed and accuracy of such judgment, models have been developed with an idea that after presentation of stimuli, an individual sequentially evaluates and accumulates information from the stimuli. When the accumulated information reaches a response criterion, the individual is assumed to make a judgment (e.g., respond with “yes” or “no”).

Unlike static models which aim to explain choice behavior alone, dynamic models intend to explain both speed and accuracy of judgment, especially trade-offs an individual has to make between speed and accuracy. It is often possible to make a correct judgment in a perceptual task, if an individual is willing to spend

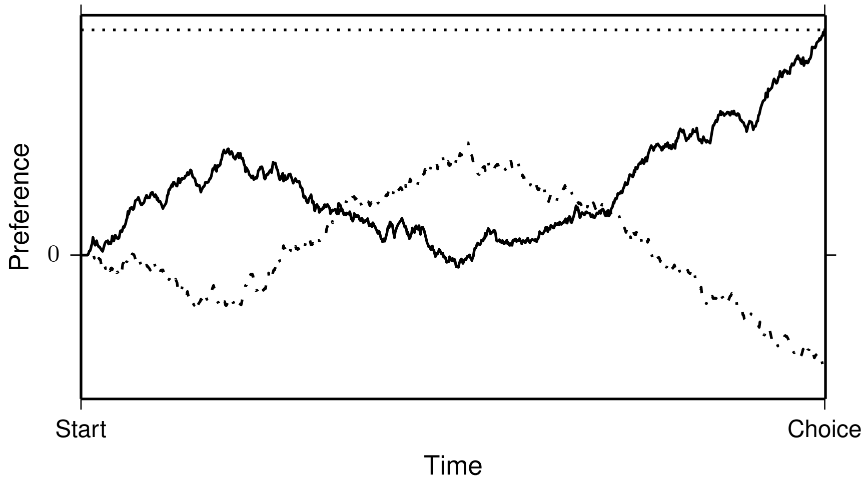


Figure 1.4: Example preference development. Two lines represent preference for two alternatives, and dotted line is a response threshold.

infinite amount of time in making a judgment. However, it typically takes much less than a second to identify a perceptual stimulus (e.g., Ratcliff, 1978), indicating a trade-off between speed and accuracy of judgment. As a result, dynamic models often provide an account for process of judgment, while static models typically do not provide such an account.

Over the last five decades, various dynamic models have been proposed and successfully explained speed-accuracy trade-offs in various perceptual judgments (e.g., Stone, 1960; Ratcliff, 1978; see Bogacz, Usher, Zhang, & McClelland, 2007 for review). Following these success, dynamic models have been extended to provide a unifying framework to explain economic choices. When applied to economic choice, a dynamic model is considered to be a model of preference development. The moment an individual faces alternatives, the individual may not have preference for an alternative over another. Rather, the individual is expected to sequentially evaluate alternatives and develop preference for each alternative.

An example of preference development is displayed in Figure 1.4. At the beginning of preference development, two alternatives have the same preference, and over time, preference is accumulated. When a preference for one alternative, represented as the solid line, reaches a threshold, the dotted line, an individual makes a choice. The nature of preference development forms a key component of dynamic models, and various models have been proposed to reflect psychologically feasible processes of preference development.

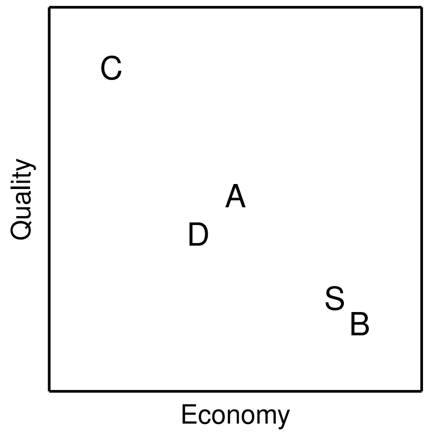


Figure 1.5: Illustration of alternatives in multi-alternative choice. Choice between Cars A and B can be affected by the presence of Car D, C, or S.

1.2.1 Decision field theory

One of the most well-known dynamic models is decision field theory (Busemeyer & Townsend, 1993). The process implemented in decision field theory is as follows. When confronted with several alternatives to choose from, an individual tries to evaluate all of the attributes associated with each alternative. Values across multiple attribute dimensions cannot be evaluated at the same time, and hence, the individual undergoes a slow and time-consuming process of evaluating, comparing, and integrating the comparisons on single attribute dimension at one time. No choice is made until the preference for one alternative becomes strong enough to guide the individual into choice. Decision field theory successfully explains various phenomena in choice between two alternatives, including those reviewed under static models (see Busemeyer & Townsend, 1993, for details).

To explain choice between three alternatives, decision field theory has been extended to multi-alternative decision field theory (MDFT; Roe, Busemeyer, & Townsend, 2001). In particular, MDFT is developed to provide an unifying explanation for the attraction, compromise, and similarity effects. These three effects document that choice can depend on the context determined by the available alternatives. An example choice between cars is illustrated in Figure 1.5. Here, each car is described in terms of two attributes, economy and quality. In this example, whether an individual chooses Car A or B can depend on the presence of Car D, C, or S in a choice set. These context effects are discussed in more detail in Chapters 5 and 6.

Importantly, the context effects suggest that preference for an alternative

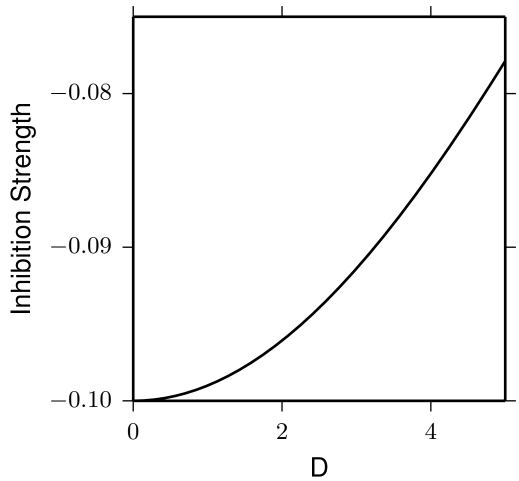


Figure 1.6: Example inhibition function: $-0.10\ \exp(-0.01\, D^2)$. Parameter values are taken from Hotaling, Busemeyer, and Li (2010).

depends on its relationships with other alternatives. Thus for a dynamic model to explain the context effects, a model needs to allow preference for an alternative to impact on preference for another alternative. In this vein, MDFT lets the developing preferences inhibit each other: increasing preference for an alternative causes preference for another alternative to decrease over time. Strength of such inhibition depends on the distance between alternatives, with more similar alternatives more strongly inhibiting each other (Tsetsos, Usher, & Chater, 2010).

To describe computational details of MDFT, I label three alternatives as A, B, and T, and denote the economy dimension as E and the quality dimension as Q. The value of Alternative A on the economy dimension is denoted as A_E and that on the quality dimension is A_Q. Preference for the three alternatives is organized in a column vector, P. The first element in this vector corresponds to preference for Alternative A, the second corresponds to preference for B, and the third corresponds to preference for T. This preference is iteratively updated as follows:

$$P(t+1) = S\, P(t) + V(t+1),$$

where S is a 3×3 feedback matrix and V is a 3×1 momentary valence vector.

In the feedback matrix, the influence of Alternative A on B is computed as:

$$-\phi_2\ \exp(-\phi_1\, D_{AB}^2).$$

Here, D_{AB} is a distance between Alternatives A and B, which is defined as a weighted sum of distance along two orthogonal vectors:

$$D_{AB} = \frac{(A_E - A_Q - B_E + B_Q)^2}{2} + \xi \frac{(A_E + A_Q - B_E - B_Q)^2}{2}.$$

Also, the self feedback is computed as $1 - \phi_2$. An example inhibition function is plotted in Figure 1.6. As a distance between alternative, D, increases, inhibition strength moves closer to 0, indicating that preference for one alternative affects the other to a less extent. This distance function is a crucial component for MDFT to explain the attraction, compromise and similarity effects (Tsetsos et al., 2010).

The momentary valence vector is computed with four matrices:

$$V(t) = C\,M\,W(t) + C\,\epsilon(t), \tag{1.5}$$

where

$$C = \begin{bmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{bmatrix},$$

$$M = \begin{bmatrix} A_E & A_Q \\ B_E & B_Q \\ T_E & T_Q \end{bmatrix},$$

and

$$\epsilon(t) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \right).$$

Here, the matrix C indicates that an attribute value of a car is evaluated in relation to the mean average values of the other cars, and ϵ indicates that the evaluation is rather noisy. As parameter value for σ increases, the evaluation becomes noisier. The attention weight W is a 2×1 vector. There is a .50 probability that

$$W(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{1.6}$$

and also a .50 probability that

$$W(t) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{1.7}$$

Equations 1.6 and 1.7 indicate which dimension is attended at given moment. When W is specified as Equation 1.6, the first dimension, the economy of cars, is attended and the three cars are evaluated according to Equation 1.5. Thus, these equations specify the process where each car is evaluated, for example, on the economy dimension at one moment, then on the quality dimension at next moment, again on the quality dimension at next moment, and so on.

At first, preference for each car is 0, and each car is iteratively evaluated, and when the highest preference reaches the threshold, the alternative is chosen. MDFT has been reported to provide a better fit to empirical choice data than probit regressions which use attribute values to predict choice (Berkowitsch, Scheibehenne, & Rieskamp, in press).

1.2.2 Comparison-grouping model

While multi-alternative decision field theory successfully explains the attraction, compromise, and similarity effects, various other dynamic models have been proposed to explain the same effects without clear indication on which model provides better explanations for actual choice behavior.

Here, I review comparison grouping model (Tsuzuki & Guo, 2004). In contrast to multi-alternative decision field theory, which assumes stochastically fluctuating attention over attribute dimensions, comparison grouping model assumes stochastically fluctuating attention over alternatives. At one moment, an individual attends two or three alternatives in a choice set and develops preference for the alternatives.

As an illustration, I again label three alternatives as A, B, and T, which are described with two attributes, E (economy) and Q (quality). The value of Alternative A on the economy dimension is denoted as A_E and that on the quality dimension is A_Q. In the comparison grouping model, each alternative and each attribute dimension iteratively develops preference. I denote preference for Alternative A as P_A. Then,

$$P_A(t+1) = P_A(t) + \Delta_A(t+1). \tag{1.8}$$

If Alternative A is not attended at time $t+1$, $\Delta_A(t+1)$ is 0, otherwise

$$\Delta_A(t+1) = \begin{cases} \zeta_A\,(1 - P_A(t)) - \lambda\, P_A(t) & \text{if } \zeta_A > 0 \\ \zeta_A\, P_A(t) - \lambda\, P_A(t) & \text{if } \zeta_A \leq 0, \end{cases} \tag{1.9}$$

where parameter λ reflects how much preference decays over time, and

$$\zeta_A = W_{A_E}\, P_E(t) + W_{Q_A}\, P_Q(t) - \tau\,(P_B(t) + P_T(t)), \tag{1.10}$$

and

$$W_{A_E} = \frac{(\ln(A_E + \mu) - \nu)}{\psi.}$$

Here, W_{A_E} determines a weight given to the economy dimension for Alternative A, which is multiplied by preference for the economy dimension, P_E. Also, parameter τ controls strength of inhibition. As with multi-alternative decision field theory, comparison grouping model allows preference for an alternative to impact on preference for another alternative. Preference for the other alternatives is updated in the same manner.

In addition, preference for attribute dimensions is updated at each iteration: P_E is updated using Equations 1.8 and 1.9, but instead of Equation 1.10, we have

$$\delta_E = W_{A_E}\, P_A(t) + W_{B_E}\, P_B(t) + W_{T_E}\, P_T(t).$$

Thus, preference for the economy dimension is updated to be an average of preferences for the alternatives, weighted by each alternative's weight given to the economy dimension. Preference for the quality dimension is updated in the similar manner. The iteration is initiated with preference for attribute dimensions and for alternatives starts with a random sample from the uniform distribution between 0.25 and 0.75. After 100 iterations, the alternative with the highest preference is chosen.

Given the process above, preference development crucially depends on probability that each alternatives is attended. If one alternative is more frequently attended, this alternative is more likely to develop preference and hence is more likely to be chosen. The probability of attention, however, is manually specified in Tsuzuki and Guo (2004), and its mathematical specification is not available in a way to apply the model to arbitrary sets of alternatives.

This lack of mathematical specification makes it difficult to test prediction from comparison grouping model, and it is not clear whether comparison grouping model provides a better explanation of the context effects than multi-alternative decision field theory. These models are tested in Chapter 5.

1.2.3 Decision by sampling

The last model to be reviewed in this chapter is decision by sampling (Stewart, Chater, & Brown, 2006). Unlike the other dynamic models reviewed above, decision by sampling is developed primarily to explain choice between two alternatives with probabilistic pay-offs and has not been extended to explain the context effects. The extension is proposed in Chapter 6.

In decision by sampling (Stewart et al., 2006), the evaluation of an alternative on a particular attribute dimension follows the rank position in a sample of attributes, and the rank position is stochastically, iteratively constructed through a series of comparisons. As an illustration, suppose an individual is making a choice between cars. The individual may attend to economy of cars at one moment and compares a car against another car in his or her memory. If the comparison favors the car in a choice set, the individual develops preference for the car. After this comparison, the decision maker may attend another attribute dimension (e.g., quality) at the next moment and makes a comparison. This iterative comparison is repeated until the preferences for the available alternatives are sufficiently different.

Here, an alternative is compared against a sample from memory on single attribute dimension at one time. Also, the preference development in decision by sampling is insensitive to the magnitude of the difference. Rather, the preference is proportional to the frequency count of the number of favorable comparisons. The ranks derived in this way replicate a number of empirical findings: for instance, the concave utility function and loss aversion (Stewart et al., 2006). The decision by sampling also provides a reasonable explanation of choice with two alternatives, especially compared against various static models (Stewart & Simpson, 2008).

Empirical support for decision by sampling is provided by several studies investigating effects of samples in memory. Ungemach, Stewart, and Reimers (2011), for example, offered customers leaving a supermarket the opportunity to choose one of two lotteries: a safe lottery with a 55% chance of winning £0.50 or a risky lottery with a 15% chance of winning £1.50. Ungemach et al. (2011)'s results show that the customers who shopped more items priced between £0.50 and £1.50 are more likely to choose the risky lottery. The prices of purchased items are likely to be available in the customers memory, and therefore, when the customers are making choice between the lotteries, the lotteries pay-offs are likely to be compared against monetary values of purchased items. When compared against values between £0.50 and £1.50, the comparison favors the risky lottery but not the safe lottery, leading to the relative preference for the risky lottery. Similar results are reported on riskless choices: for example, a choice between job applicants (Mellers & Cooke, 1994) and apartments (Cooke & Mellers, 1998). These studies show that, for example, the individuals who frequently saw monthly rents between $350 and $400 find the rent of $350 much more attractive than $400.

1.3 Plan of thesis

The static models, reviewed above, are examined in Chapter 2. Chapter 2 develops a non-parametric method for estimating utility over the wide range of alternatives. The estimates are compared against expected utility theory, cumulative prospect theory, and the transfer of attention exchange model. Extensions of expected utility theory and cumulative prospect theory are examined in Chapter 3, which investigates impacts of choice set-size and also of information presentation formats. These impacts are further examined in Chapter 4. Then, Chapter 5 critically evaluates dynamic models of multi-alternative choice with process-tracing data, and the process-tracing data are further leveraged to propose an extension of decision by sampling in Chapter 6. Then Chapter 7 summarizes results from Chapters 2 to 6 and concludes this thesis.

Chapter 2

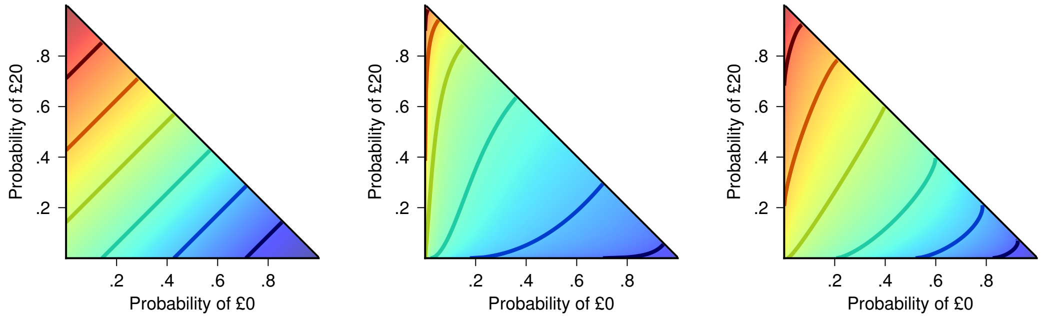
Non-parametric estimation of the individual's utility map

2.1 Background

Understanding how people trade off risk and reward is a fundamental goal of behavioral economics. The most common approach to modeling how people make decisions between risky alternatives is based on the idea of utility: individuals integrate their subjective probability of reward with their subjective value of the reward to produce a single value, their utility that describes how well the alternative is preferred. The utilities of the alternatives are then compared and the alternative with the highest utility is most often chosen.

The normative calculation of utility that maximizes long-term gain is to multiply the probability with the subjective value of the associated outcome. For an illustration, suppose an individual is considering a choice alternative with three possible outcomes: £20, £10, and £0. This particular alternative has a 20% probability for £20, 40% for £10, and 40% for £0. Then, the expected utility is $20\% \times v(£20) + 40\% \times v(£10) + 40\% \times v(£0)$, where v is the function to map the monetary value to the subjective value.

However, previous research has demonstrated that an individual's choice frequently deviates from the predictions of expected utility theory (for review, Schoemaker, 1982). To explain the deviations, descriptive models of how risk and reward are integrated have been developed (for review, Starmer, 2000). A common and useful way to visualize the predictions of these models is to look at the indifference lines, which connect choice alternatives of equal utility, over a probability triangle (Camerer, 1989). The probability triangle is a two-dimensional space which maps



(a) The expected utility theory with the identity value function

(b) The cumulative prospect theory with parameters $\alpha = 0.88$ and $\gamma = 0.52$

(c) The transfer of attention exchange model with parameters $\beta = 1$, $\gamma = 0.7$, and $\delta = 1$

Figure 2.1: Predicted utility map over a probability triangle. A point in this triangle represent an alternative, whose potential outcomes are £20, £10, and £0. Probability of £20 is shown on the vertical axis, and probability of £0 is shown on the horizontal axis.

alternatives with varying probabilities for the same set of three potential outcomes. Throughout this chapter, we use £20, £10, and £0 as the potential outcomes from a choice alternative.

Figure 2.1 displays the predicted utility maps from three of the most well-known models of risky choice: expected utility theory, cumulative prospect theory (Tversky & Kahneman, 1992) and transfer of attention exchange (TAX) model (Birnbaum, 2008). The differences between the models can be seen in the shapes of the indifference lines. Expected utility theory predicts indifference lines that are parallel and straight. Cumulative prospect theory predicts lines that are concave where the probability of the best outcome is larger than that of the worst outcome, but convex lines where the probability of the best outcome is less than that of the worst. The TAX model predicts indifference lines that are convex throughout the triangle.

The usual experimental practice is to investigate choices in regions of the triangle where models most differ from each other (e.g., Wu & Gonzalez, 1998). When the models are tested in this way, the best model may not predict choices away from the diagnostic regions well. For instance, Harless (1992) reports that the cumulative prospect utility explains choice better than the expected utility theory only in the boundary region of the triangle.

To assess models using the entire region of the triangle, we develop a non-parametric method to estimate entire utility maps, an extension of Markov chain Monte Carlo (MCMC) with People (Sanborn, Griffiths, & Shiffrin, 2010). We have modified MCMC with People to investigate regions of the probability triangle where the choice alternatives are less preferred. The new method is tested in a simulation to show that it can deliver useful results within a reasonable number of trials. We then estimate utility maps from human data and determine which model fits best. Finally, we discuss the results and future applications for this approach.

2.1.1 Markov chain Monte Carlo with People

Markov chain Monte Carlo (MCMC) is a common method for drawing samples from a distribution. It has been widely used to draw probabilistic inference especially when solving the exact function of interest is computationally difficult (Neal, 1993). Samples drawn with MCMC are typically used to make an inference on the distribution, and here, we use samples to infer the shape of utility map over the triangle.

MCMC begins in a start state z. Then a sample z' is first drawn from the proposal distribution q, then z' is evaluated with the function of interest, π to determine whether to accept z' as a new state or discard it and retain the current state z. The sequence of accepted samples forms a Markov chain, and after this Markov chain converges, accepted samples can be regarded as samples from π distribution. To ensure that the Markov chain converges to π, detailed balance needs to be satisfied

$$\pi(z)\, q(z'|z)\, A(z', z) = \pi(z')\, q(z|z')\, A(z, z'), \tag{2.1}$$

where $q(z'|z)$ is the probability of drawing z' when the current state is z and $A(z', z)$ is the probability of accepting proposal z' over the current state z.

Throughout the chapter, we assume a symmetric distribution for q, $q(z'|z) = q(z|z')$, so Equation 2.1 becomes

$$\pi(z)\, A(z', z) = \pi(z')\, A(z, z'). \tag{2.2}$$

Detailed balance can be satisfied by carefully designing the acceptance function A. The most commonly used function is the Metropolis acceptance function (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953), but the Boltzmann acceptance function (Flinn & McManus, 1961) is of interest here:

$$A(z', z) = \frac{\pi(z')}{\pi(z) + \pi(z').}$$

If an individual is asked to make a choice between alternatives z' and z, then the Boltzmann acceptance function can model that individual's choice. This is because the Boltzmann function is equivalent to Luce's choice rule (Luce, 1959), which has been frequently used to model risky choice (e.g., Loomes & Sugden, 1998; Blavatskyy & Pogrebna, 2010). As a result, by sequentially presenting pairs of choice alternatives to an individual (where the new alternative z' is selected by the computer), the collection of choice alternatives chosen by the individual can be treated as samples from the probability distribution whose density is proportional to the individual's utility.

2.1.2 Extending MCMC with People

However, samples from the individual's utility distribution does not necessarily serve to estimate the shape of the utility map: pilot work confirms that all of the samples will be concentrated around the most favorable alternative (100% probability of £20 in the triangle), leaving the rest of the utility map unexplored. To enable the reasonable estimation of the utility map, the Markov chain has to travel better around the triangular space.

For this purpose, we implement a latent agent in the experimental program. This latent agent makes an independent choice between the same alternatives as participant, and only when the agent and participant both select the new choice alternative, the new alternative becomes the new state. Otherwise, the current state remains the same and another alternative is generated from the proposal distribution.

When the agent is implemented in this way, the acceptance function becomes a joint function of the participant's and the agent's choices. Specifically, the acceptance function is defined as

$$A^*(z', z) = \frac{f(z')}{f(z) + f(z')} \frac{g(z')}{g(z) + g(z'),}$$

where f is the utility function for participant and g is the agent's utility function. Here, both participant and the agent follow the Boltzmann acceptance function. Then Equation 2.2 becomes

$$f(z)\, g(z)\, A^*(z', z) = f(z')\, g(z')\, A^*(z, z').$$

With the implementation of the agent, the trajectory of the Markov states depends on both participant's and the latent agent's choices. If the agent's utility follows an inverse of optimal choices (i.e., the utility is the lowest at the top corner of the triangle), the Markov chain would be pushed away from that region. Thus with this extended method, the stationary distribution of the Markov chain is the joint utility function of the participant and the agent, $f\,g$. Participant's utility map can subsequently be recovered by dividing the joint utility by the latent agent's known utility.

2.2 Simulation

To demonstrate that the developed method can estimate a participant's utility map within a reasonable number of trials, a simulation was run. The simulation used the two of the utility functions in Figure 2.1: g was set to the inverse of expected utility theory, and f was cumulative prospect theory. The proposal distribution, q, was uniform over the triangular space. The possible outcomes were fixed to be £20, £10 and £0.

With these functions, a choice trial was simulated as follows. First, the agent used the g function to evaluate each alternative and uses the Boltzmann acceptance function to select between the current state and the proposed alternative. If the agent preferred the current state over the proposed alternative, another alternative was sampled from the proposal distribution. If the latent agent chose the new alternative over the current state, the virtual participant with f function then made a choice between the same two alternatives.

Although the agent and the virtual participant could have made a choice at the same time over the same two alternatives, we had the agent decide first: if the agent does not select the new alternative, the previous state remains the state regardless of the choice the participant makes. This reduces the number of choices the participant must make.

Each simulation consisted of three chains: one chain started with the Markov state of 60% of £20, 20% of £10 and 20% of £0. Another chain started with the state of 20% of £20, 60% of £10 and 20% of £0 as the starting state. The final chain started with 20% of £20, 20% of £10 and 60% of £0.

The first 100 trials were considered to be trials before convergence of the Markov chain (burn-in period) and were discarded from each chain. The remaining samples from the three chains were pooled and smoothed by kernel density estimation. Because of the triangular boundary of the estimation space, it is actually quite difficult to produce unbiased indifference curves. We chose to use a Dirichlet kernel, an extension of the Beta kernel (Chen, 1999) to the triangular space, because it

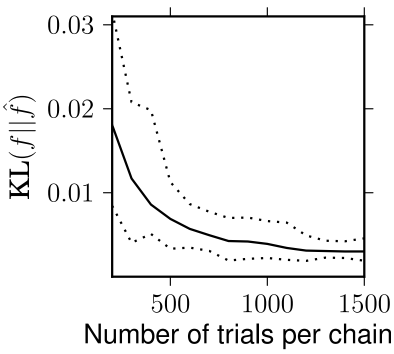


Figure 2.2: The KL divergences between f and $\hat{f}$ for various numbers of trials. The solid line represents the mean measurement of the 10 simulation runs, and the dotted lines are maximum and minimum values archived in the simulations.

produced less bias than other alternatives. The Dirichlet kernel is defined as

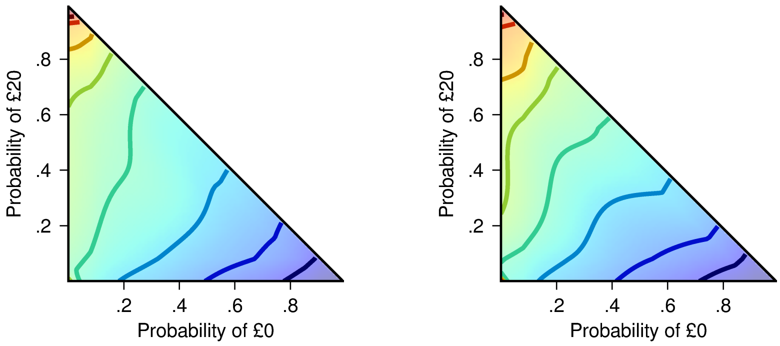
$$\hat{f}(x)\, g(x) = \sum_i Dir(z_i|\alpha_1, \alpha_2, \alpha_3),$$

where z_i is the ith state in the Markov chain, x is a vector of probabilities for three outcomes, and α_j is $x_j/min(h, x_j, 1-x_j)$. The kernel width h was set to 0.09. This smoothed joint distribution is then divided by g to derive the estimation $\hat{f}$.

To assess the similarity between f and $\hat{f}$, we computed Kullback–Leibler (KL; denoted as $\mathrm{KL}(f||\hat{f})$) divergence (Kullback & Leibler, 1951), which measures how much information is lost in the estimation process. The KL divergences for different sample sizes are plotted in Figure 2.2. This figure illustrates that the estimation shows the increasingly smaller divergence within the first few hundred trials. The estimation becomes reasonably accurate on average after 700–800 trials.

The two panels of Figure 2.3 display the estimations after 1,000 trials. The estimation with the smallest KL divergence among the 10 simulation runs is in the left panel, and the right panel show the estimation with the largest KL divergence. Both panels show the key properties of the cumulative prospect theory: the estimated maps display the concave indifference lines where the probability of £20 is greater than the probability of £0, and the indifference lines are convex in the other area. Also, the indifference lines show fanning-out property from the lower left corner toward the diagonal boundary.

Thus, the simulation demonstrated that the proposed method with the Dirichlet kernel density estimation can recover the key characteristic of the utility map using a reasonable number of samples.



(a) The estimation with the smallest KL divergence ($\text{KL}(f||\hat{f}) = 0.002$)

(b) The estimation with the largest KL divergence ($\text{KL}(f||\hat{f}) = 0.007$)

Figure 2.3: Estimation of the cumulative prospect theory with 1,000 trials

2.3 Experiment

2.3.1 Method

Participant

Ten participants were recruited through the subject panel at the University of Warwick. One participant did not complete the experiment, leaving nine (five male and four female) participants. Their age ranged from 19 to 30 with a mean of 22.9.

Procedure

The experimental procedure closely followed that of the simulation. Three chains with the same start states were run interleaved until participants had made 1,000 choices per chain. In each trial, the latent agent made a decision first, and a new alternative was drawn from the uniform distribution over the entire triangle until the agent chose the new alternative. The latent agent's utility function was set to be the prediction from the expected utility theory raised to the power of -8, which was enough to ensure coverage of the map in pilot work.

In addition, 50 catch trials were inserted to the experiment, so that we could assess whether participants were engaged in the task. In each catch trial, one alternative had larger probabilities for both £20 and £10. If a participant was not engaged with the task and randomly making choices, it is expected that he or she would occasionally not select the non-dominant alternative.

The experiment presented a choice alternative as a pie chart with three slices. Each slice represented one possible outcome, and the size of the slice was proportional to the probability of the outcome. Participants were forced to log out from the online experiment and take a break after spending one hour on it. After the minimum break of three hours, participants were allowed to log in again and resume the experiment.

The choices participant made were incentivized: we invited participants to the lab when participants completed the experiment. At the lab, we randomly selected one trial from the experiment and played the selected alternative for real. Participants were paid what they earned from the play.

2.3.2 Result and Discussion

All the participants selected the dominant alternative in all of the catch trials, which was evidence that all participants understood and were engaged in the task.

Utility maps were estimated as in the simulation study. All participants show a sharp peak at the top corner of the triangle in the estimated maps. The sharp peak makes it difficult to see the shape of the map, and thus for illustration purposes, we spaced out the indifference lines by taking the natural logarithm of the estimation. As a result, small differences in utility are exaggerated, but the shapes of the indifference lines are not affected. The resulting maps are displayed in Figure 2.4. Each panel in the figure corresponds to one participant's map.

The estimated maps show the steep indifference lines, especially where the probability of £0 is small. The steep lines indicate aversion to the worst outcome (c.f., Brandstatter, Gigerenzer, & Hertwig, 2006), where the increment in probability for the worst outcome needs to be compensated with a larger increment in probability for the most desirable outcome. The steepness tends to be lessened near the lower right corner of the triangle. As a result, for Participants A, D and H in particular, the indifference lines show the fanning-out property. The fanning-out suggests that participants more willingly accept an increment in probability for the worst outcome when the probability is already large. The fanning-out is consistent with the prediction from the cumulative prospect theory.

The estimated maps also shows the convex indifference lines, rather than the concave, throughout the triangle. The convexity makes the estimated maps appear similar to the predicted utility map from the transfer of attention exchange (TAX) model (Figure 2.1). To assess the similarity between the estimated maps and the model predictions, we computed the KL divergence between the estimated maps and the models. As $\hat{f}$ is a non-parametric map, the divergence is analytically intractable.

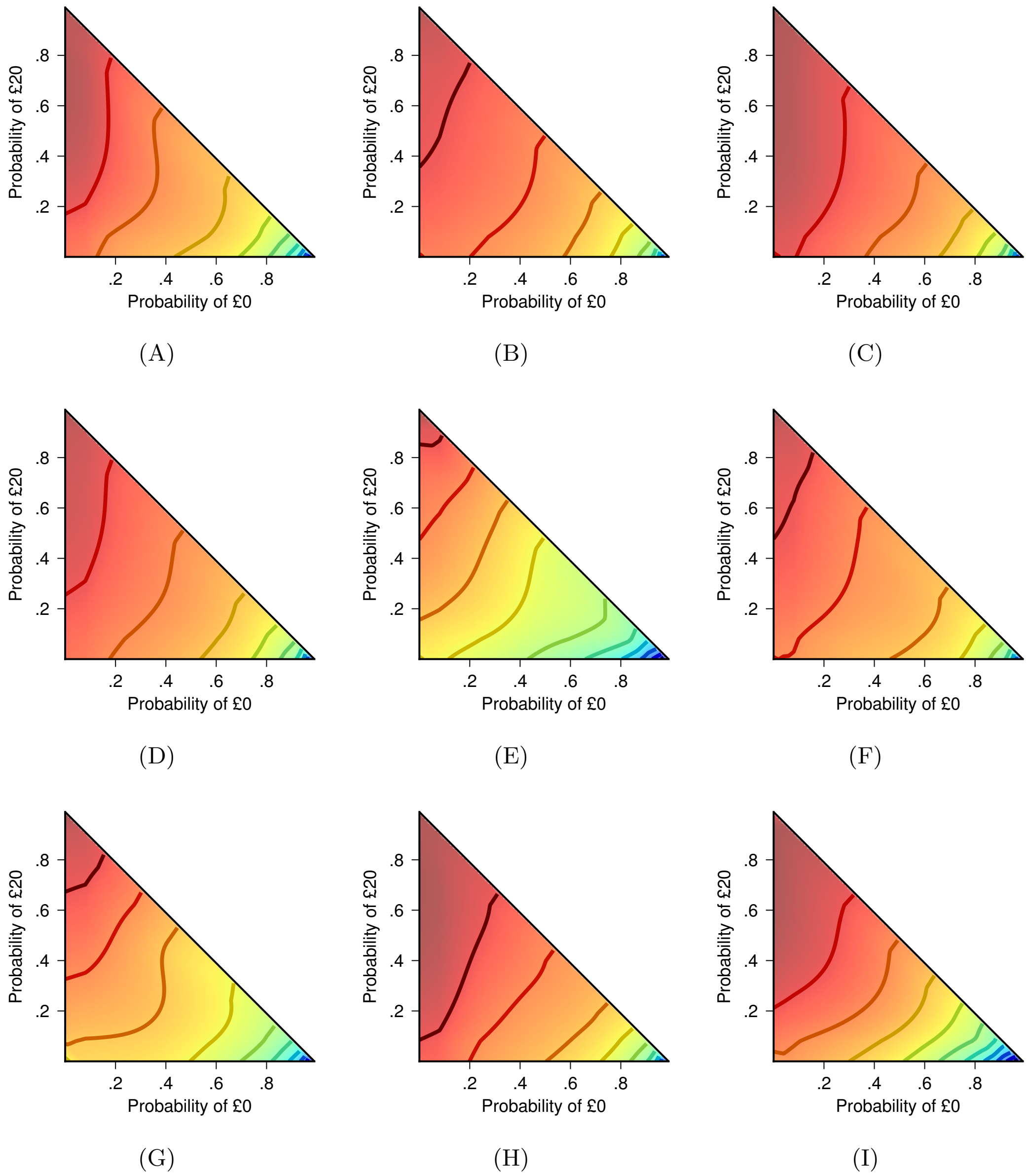


Figure 2.4: $ln(\hat{f})$: logged estimation of utility map. Each panel represents utility map for one participant. Solid line represents indifference line.

Thus we approximate the divergence with a Monte Carlo calculation:

$$\begin{aligned}
\mathrm{KL}(\hat{f}||\text{model}) &= \int p(u|\hat{f})\{\ln(p(u|\hat{f})) - \ln(p(u|\text{model}))\}\, du \\
&\approx \frac{1}{n}\sum_{i=1}^{n}\{ln(p(\hat{u}_i|\hat{f})) - ln(p(\hat{u}_i|\text{model}))\} \\
&\approx \frac{1}{n}\{ln(p(\hat{U}|\hat{f})) - ln(p(\hat{U}|\text{model}))\},
\end{aligned}$$

where $\hat{u}_i$ is the ith element in a vector $\hat{U}$ of n independent samples from $\hat{f}$. We used 10^4 as n. The marginal likelihood is computed as follows:

$$p(\hat{U}|\text{model}) = \int p(\hat{U}|\theta,\ \text{model})\ p(\theta|\text{model})\ d\theta.$$

Here θ is the model parameters, and $p(\theta|\text{model})$ is the prior probability of θ. The prior was the uniform distribution between 0 and 2 for all the parameters. This marginal likelihood penalizes unnecessary model complexity (Myung & Pitt, 1997), and thus the KL divergence computed with the marginal likelihood also penalizes unnecessary model complexity.

To allow the predicted maps to adjust the spacing between the indifference lines, the parameter vector θ includes one additional parameter used as an exponent for the predicted utility. If this parameter value is greater than 1, the predicted map produce the sharper peak. The prior for this exponent parameter is the uniform between 0 and 10. Also, for the expected utility theory, we used the power law value function: $v(s) = s^{\alpha}$.

The approximated KL divergence is mean-averaged over participants: the means are 0.80 (SE = 0.10) for the expected utility theory, 0.85 (SE = 0.08) for the cumulative prospect theory, and 1.09 (SE = 0.11) for the TAX model. These mean divergences are below KL divergence of 2.06 (SE = 0.22) between the estimated map and the uniform map. Thus, all three theories provide better predictions than the uniform map, assigning the same utility to all alternatives. The divergence from the expected utility theory is smallest for seven out of nine maps (Panels A, C, D, E, G, H, and I in Figure 2.4), and for the remaining maps, the divergence from the cumulative prospect theory is smallest. The divergence from the TAX model is largest for all the maps.

However, the TAX model achieves largest maximum likelihood for the four maps (Panels C, G, H, and I). Thus, it is possible that with larger n, the TAX model archives larger marginal likelihood for these four maps, resulting in smaller divergence. Nonetheless, for the remaining five maps, the maximum likelihood is largest for the cumulative prospect theory.

2.4 General Discussion

We have estimated non-parametric utility maps over the probability triangle. The estimated maps indicates that the sensitivity to probability depends on the associated outcome and also the magnitude of the probability. The probability of the worst outcome is more heavily weighted than that of the better outcome. Also, the sensitivity to the increment of the probability diminishes as the probability increases, but this diminishing sensitivity is applied more readily for the probability of the worst outcome.

The curvature of the indifference lines in the estimated maps appears similar to the prediction from the attention exchange (TAX) model. However, the steepness of the indifference lines is more in line with the cumulative prospect theory. The model comparison indicates that five out of the nine maps are closer to the cumulative prospect theory and the remaining four maps are closer to the TAX model.

While the TAX model in Figure 2.1 predicts the convex indifference lines throughout the triangle, the estimated indifference lines tend to flatten out and form straight parallel lines toward the lower right corner of the triangle. In contrast, the indifference lines do not appear flattening out toward the upper left corner. This varying curvature implies that the performance of the model can differ in the corner of the triangle, and further indicates that the choice preferences need to be examined over the broader region of the triangle.

Then, it is of theoretical interest to identify choice alternatives where the model prediction differ from the individuals' choice behavior. To this end, the estimation method that we have developed can be further extended. For instance, by setting the latent agent's utility to the inverse of the TAX model, the MCMC chain converges to the distribution whose density is proportional to the individual's utility divided by the TAX model prediction. The condensed area in this joint utility distribution is where the TAX model underpredicts the utility, and the thin area is where the TAX model overpredicts the utility.

Also, the estimated maps could be further leveraged for data-driven analysis. For instance, similarity between the maps could be quantified with Kullback-Leibler divergence, and the maps could be classified into several categories. Such classification might help identifying cluster of individuals with distinctively different pattern of utility map.

To conclude, we have developed the method for estimating the utility map and comparing models, and shown that model comparison can benefit from considering the broader range of alternatives. The developed method can be further leveraged in future study. Utility, however, may be unstable and sensitive to the choice context, as the results from the following chapters indicate.

Chapter 3

Set-size induced risk-amplification: Experience-based decisions in large set-sizes favor riskier alternatives

3.1 Background

Over the past decade, research with two-alternative environments has led to the claim that in the sampling paradigm, where a choice is made after sampling a series of sample pay-offs (such as \$0, \$0, \$0, \$9, and \$0 from one alternative), individuals make a choice as if they under-weight small probabilities (Hertwig, Barron, Weber, & Erev, 2004). This under-weighting has been juxtaposed against over-weighting of small probabilities in decisions from description (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), where pay-offs and their probabilities are described (e.g., \$9 with a 10% probability, otherwise nothing). This difference in the weighting of small probabilities — termed the description-experience gap (Hertwig & Erev, 2009) — has been rigorously examined and consistently confirmed in two-alternative environments (e.g., Ungemach, Chater, & Stewart, 2009; Lajarraga, Hertwig, & Gonzalez, 2012; Gonzalez & Dutt, 2011; Newell & Rakow, 2007; Erev et al., 2010; Gottlieb, Weiss, & Chapman, 2007; Abdellaoui, L'Haridon, & Paraschiv, 2011; Hilbig & Gloeckner, 2011).

The description-experience gap has been used to explain a variety of phenomena related to decision making outside the laboratory, including those involving the financial crisis (Hertwig & Erev, 2009) and perceived terrorist threats (Yechiam, Barron, & Erev, 2005). As an illustration, suppose an individual is assessing how likely he is to lose weight from a specific diet. One method of assessment is to recall other individuals who have tried the diet (Tversky & Kahneman, 1973; Galesic, Olsson, & Rieskamp, 2012). If the diet rarely leads to weight loss, then the probability of weight loss may be under-weighted and the individual may infer that the diet is a waste of time.

However, because the number of alternatives is often more than two outside the laboratory, generalization of the description-experience gap requires that the under-weighting and subsequent choice be independent of the number of alternatives. Continuing the above example and generalizing from the description-experience gap, the individual should under-weight the probability of weight loss and choose not to try dieting regardless of the number of diets he considers.

Choices are, however, often influenced by the context provided by choice sets (e.g., Huber, Payne, & Puto, 1982; Simonson, 1989; Tversky, 1972), and prior empirical evidence suggests that a choice can change systematically with a growth in set-sizes. Hills, Noguchi, and Gibbert (2013), for example, found that a larger and more diverse set leads individuals to take more samples overall but to sample fewer times per alternative and to subsequently choose alternatives which delivered a larger pay-off.

Further, when individuals can observe foregone pay-offs from the alternatives they did not choose, a larger set tends to lead the individuals to choose the alternative that, most recently, delivered the largest pay-off (Ert & Erev, 2007). Ert and Erev (2007) also show that choices are sensitive to the difference between a large pay-off observed prior to a choice and the subsequent pay-off they earned from the choice. If the difference is large, individuals eventually stop choosing the alternative that most recently delivered a large pay-off. Thus through feedback individuals learn that the small number of pay-offs they observed are not representative. In most studies with the sampling paradigm, however, participants do not receive feedback immediately after making a choice (e.g., Hertwig et al., 2004), and thus individuals are likely to keep choosing an alternative which delivered a large sample pay-off.

Based on the above work, we hypothesized that in the sampling paradigm, a safer alternative would be more often chosen in a small set, but a riskier alternative with a rare but large pay-off would be more likely to be chosen in a large set in the gains domain, where a pay-off is either zero or positive. This amplified risk-taking in the sampling paradigm should eventually diminish the description-experience gap, as we do not expect risk-taking to be amplified in the description paradigm. Unlike in the sampling paradigm, alternatives' risk is apparent in the description paradigm, and therefore, individuals could treat a choice in large sets as a choice between categories of riskier ans safer alternatives, making their choices invariant to set-sizes with respect to risk.

For the loss domain, where a pay-off is either zero or negative, the largest possible pay-off is often zero, and both riskier and safer alternatives can deliver zero pay-offs. Here, riskier alternatives deliver zero pay-offs more often, but riskier alternatives also occasionally deliver large negative pay-offs. If a growth in set-sizes facilitates choice for alternatives with larger pay-offs, risk-taking should not be influenced in the loss domain.

In addition to the investigation into the influence of set-sizes on a choice, the present study aims to extend previous studies by examining whether amplified risk-taking is explained by sampling error. Although the description-experience gap can be potentially attributed to a variety of factors (e.g., Hertwig & Erev, 2009; Hills & Hertwig, 2010), previous studies with two-alternative environments have demonstrated that risk-taking in the sampling paradigm is largely due to sampling error (e.g., Fox & Hadar, 2006; Ungemach et al., 2009; Hertwig & Pleskac, 2010). Fox and Hadar (2006), for example, show that individuals tend not to sample enough to observe large non-zero pay-offs from a riskier alternative. As a result, these individuals make a choice as if a pay-off from the riskier alternative can only be zero. Also, Ungemach et al. (2009) show that when sampling error is minimized by forcing individuals to sample a certain number of times from each alternative, a riskier alternative is chosen more often in the gain domain.

Thus, previous studies show that sampling error in a small set tends to promote a choice for a safer alternative in the gain domain. With a growth in set-sizes, sampling error may promote a choice for a riskier alternative in the gains domain, as at least one riskier alternative is more likely to deliver a large pay-off in a large set than in a small set.

In what follows, we first describe two experiments which examined the influence of set-sizes on risk-taking and sampling error. We then examine how the number of samples and set-sizes relate to sampling error.

3.2 Method

In Experiments 1 and 2, we manipulated set-sizes in both description and sampling paradigms, for both gain and loss domains, and for conditions in which the expected pay-off of an alternative was positively or negatively associated with the probability of pay-off. Both experiments employed a 2 (between-participant, set-size: small or large) $\times$ 2 (between-participant, paradigm: description or sampling) $\times$ 2 (within-participant, domain: gain or loss) design.

Experiment 1 was conducted on-line and the pay-off was in American dollars. Experiment 2 was carried out in a laboratory and the pay-off was in British pounds. The two experiments differed in the relationship between risk and expected pay-off to ensure that the results do not depend on the particular structure of the pay-offs. Additionally, the experiments differed in how the participation fee was calculated (explained below).

3.2.1 Participants

In Experiment 1, 131 participants (73 males, 56 females and 2 unspecified) were recruited through Mechanical Turk (`http://www.mturk.com`). Their age ranged from 18 to 69 with a mean of 30.63. In Experiment 2, 101 students (57 males and 44 females) were recruited through the participant panel at the University of Warwick. Their age ranged from 18 to 52 with a mean of 22.7. We decided in advance of collecting the data to test exactly 100 participants for both experiments, but over-recruited due to technical reasons.

3.2.2 Apparatus

The alternatives were independently and randomly generated for each trial for each participant. Half of the alternatives within a trial (1 alternative in a small set, and 16 alternatives in a large set) were *safer* alternatives, and the other half were *riskier*. Each alternative delivers a non-zero pay-off with a certain probability and otherwise nothing. Safer and riskier alternatives are characterized by different probabilities of non-zero pay-off.

Within a trial, all the safer alternatives had the same expected pay-off — a random draw from a uniform distribution between 0.50 and 1.00 for the gain domain and between -0.50 and -1.00 for the loss domain. Thus in a particular trial in the gain domain, for example, all the safer alternatives may have the expected pay-off of 0.83.

Then, the expected pay-off for the safer alternatives was multiplied by 0.9 in Experiment 1 and 1.1 in Experiment 2, to derive the expected pay-off for the riskier alternatives. If the safer alternatives in one trial had the expected pay-off of 0.83, for instance, all the riskier alternatives within the same trial had the expected pay-off of $0.83 \times 0.9 = 0.747$ in Experiment 1 and $0.83 \times 1.1 = 0.913$ in Experiment 2.

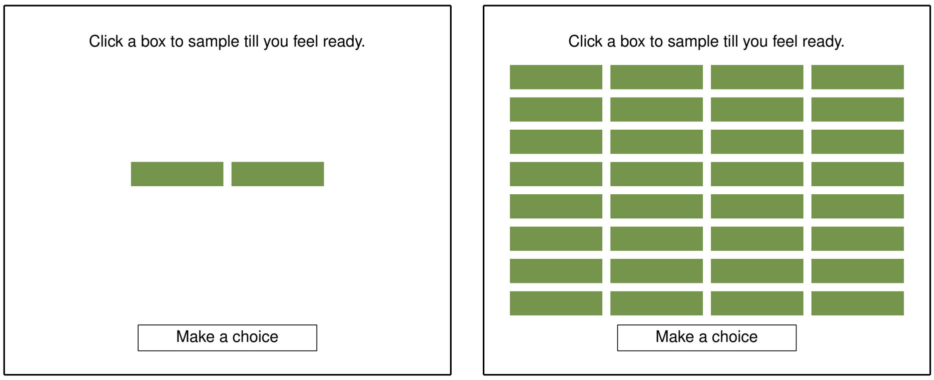
Then, each alternative was independently assigned with a probability of non-zero pay-off. For a safer alternative, a probability of non-zero pay-off was a random number drawn from a uniform distribution between 0.800 and 0.995. Then, the expected pay-off for an alternative was divided by the probability of non-zero pay-off to derive the non-zero pay-off amount. Thus, if a safer alternative had an expected pay-off of 0.83 and a .943 probability of non-zero pay-off, the amount of the non-zero pay-off is $0.83/.943 = 0.880$. Then the probability and the amount of pay-off were rounded to the nearest two decimal points. Continuing the above example, the alternative has a .94 probability of 0.88 pay-off and a $1 - .94 = .06$ probability of 0.00 pay-off.

Similarly for each of the riskier alternatives, a probability of non-zero pay-off was a random number drawn from a uniform distribution between 0.005 and 0.200. This range of probability follows Hertwig et al. (2004)'s definition of rare events. The amount of non-zero pay-off was then derived with the same procedure as described above. Please note that pay-offs from a riskier alternative tends to have a higher variance than a pay-off from a safer alternative. By eliminating extremely rare or certain pay-offs (with probability less than .005 or greater than .995), we aimed to reduce the possibility that participants would infer that non-zero pay-offs were impossible or certain.

Each alternative was presented as a box on a screen, and a set of alternatives was presented to participants as an array of boxes. An example screen-shot is given in Figure 3.1. The left panel in this figure illustrates a small set with 2 alternatives, and the right panel illustrates a large set with 32 alternatives. Locations of alternatives were randomized independently for each trial for each participant. Thus, the top far-left box may be a riskier alternative on one trial and a safer alternative on another trial.

3.2.3 Procedure

Participants were instructed that their payment would depend on their choices during the experiment. The two experiments asked each participant to make six choices, three involving gains and three involving losses. The gain and loss trials were in-



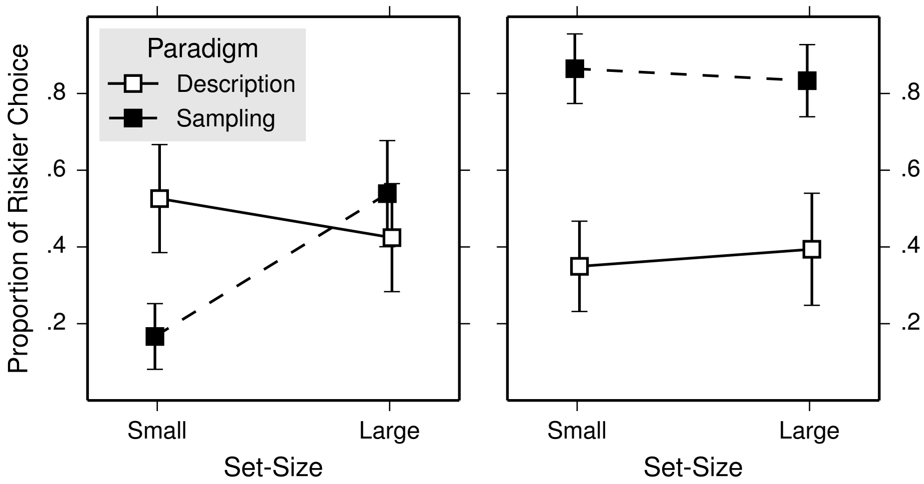
(a) Small Set (b) Large Set

Figure 3.1: Example screen-shots.

terleaved and presented in a random order. At each trial, participants saw either 2 or 32 alternatives and were allowed to draw samples from the alternatives as many times as they wanted before choosing one of the alternatives. Every time a sample was drawn, information about the alternative was presented for 500ms. In the description paradigm, the information displayed the probability and the amount of non-zero pay-off (e.g., 94%, \$0.88). In the sampling paradigm, the information presented was a sample, randomly drawn with replacement from the pay-off distribution associated with that alternative. For example, about 94 out of 100 samples from an alternative with a 94% probability of \$0.88 were \$0.88 and the other 6 were \$0.00. Participants did not learn about the pay-offs from their choices until the end of the experiment. In Experiment 1, the pay-offs from the six choices were summed and added to the base fee of \$1.00. After adding the base fee, the fee ranged from \$0.00 to \$14.57, with a mean of \$1.80. In Experiment 2, the pay-off from one choice was randomly selected and added to the base fee of £4.00. After adding the base fee, the fee ranged from £0.00 to £8.00, with a mean of £4.08.

3.3 Results

Analyses were confined to trials where participants sampled at least two alternatives and chose an alternative they had sampled at least once. This represented 686 out of 786 choices (= 131 participants × 6 choices) in Experiment 1, and 555 out of 606 choices (= 101 participants × 6 trials) in Experiment 2.



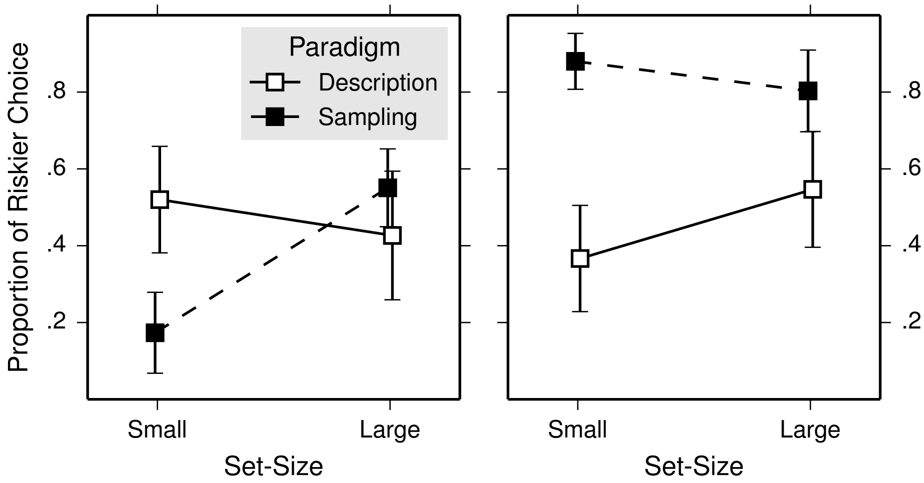
(a) Gain Domain (b) Loss Domain

Figure 3.2: Proportion of choices for riskier alternatives in Experiment 1. Error bars are 95% confidence interval.

3.3.1 Risk-taking

For each participant, we calculated the proportion of choices for riskier alternatives, as in Figures 3.2 and 3.3. These figures show that both experiments replicated the description-experience gap in the small set size for both gain and loss domains. In the gain domain, a riskier alternative was more frequently chosen in the description paradigm than in the sampling paradigm. Alternatively in the loss domain, a riskier alternative was more frequently chosen in the sampling paradigm than in the description paradigm. These patterns are consistent with the description-experience gap. However when a set-size is large, the description-experience gap disappears in the gains domain, but not in the loss domain.

Choices for riskier alternatives were examined with mixed-effect logistic regressions, which include by-participant intercepts and slopes as random factors. Model fits indicate a significant three-way interaction between the domain, the paradigm, and the set-size (Experiment 1: $\chi^2(1) = 6.71$, $\beta = -3.79$, $p < .01$; Experiment 2: $\chi^2(1) = 11.70$, $\beta = -4.11$, $p < .001$). For the gain domain, the effect of paradigm differs between the set-sizes (Experiment 1: $\chi^2(1) = 13.22$, $\beta = 3.08$, $p < .001$; Experiment 2: $\chi^2(1) = 11.84$, $\beta = 2.43$, $p < .001$). Specifically, with a growth in set-sizes, a riskier alternative is more frequently chosen in the sampling paradigm (Experiment 1: $\chi^2(1) = 19.24$, $\beta = 2.21$, $p < .001$; Experiment 2: $\chi^2(1) = 18.90$, $\beta = 1.69$, $p < .001$) but not in the description paradigm ($ps > .39$). This amplified risk-taking eliminates the description-experience gap in the large set.



(a) Gain Domain (b) Loss Domain

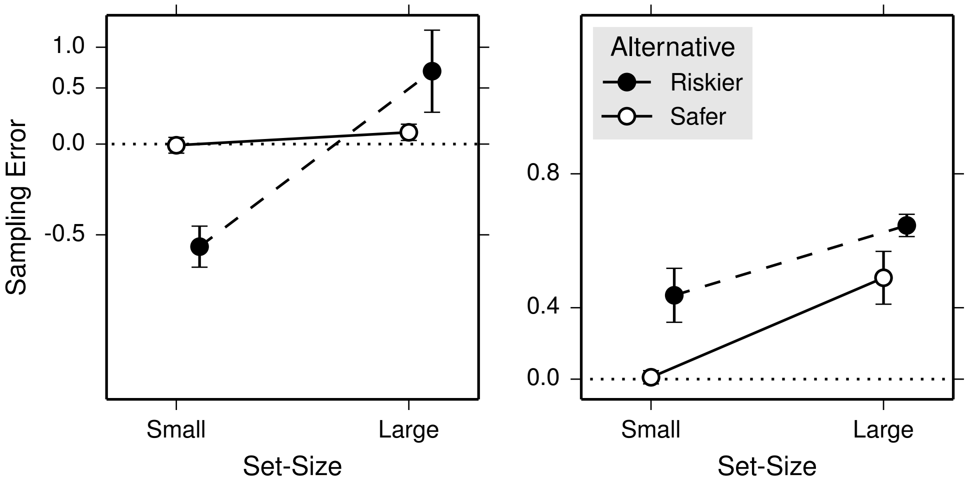
Figure 3.3: Proportion of choices for riskier alternatives in Experiment 2. Error bars are 95% confidence interval.

For the loss domain, a growth in set-sizes does not significantly influence choices for riskier alternatives ($ps > .07$). Here, the description-experience gap persists, providing reassuring support for prior generalizations from the two-alternative environments in the loss domain (e.g., Yechiam et al., 2005; Hertwig & Erev, 2009).

To show that the elimination of the description-experience gap in the gain domain is not due to a greater tendency for random choice in the large set-size, we computed the mean sample pay-offs participants observed for each alternative. Using this mean pay-off, the sampled alternatives were ranked in a descending order within each trial. Ranks were then normalized as follows:

$$\text{relative rank} = \frac{\text{rank} - 1}{n - 1},$$

where n is the number of unique alternatives sampled. Random choice is indicated by a mean relative rank of the chosen alternative close to .50. The mean relative rank of the chosen alternative in the sampling paradigm for the large set-size was .85 (SE = 0.02) in Experiment 1 and .90 (SE = 0.02) in Experiment 2. These were both significantly higher than .50 (using a mixed-effect linear regression with maximal random effects on the logit-transformed ranks, Experiment 1: $t(325) = 10.40$, $p < .001$; Experiment 2: $t(260) = 14.79$, $p < .001$), indicating that participants consistently chose higher ranked alternatives among those they sampled and thus were *not* choosing randomly.



(a) Gain Domain

(b) Loss Domain

Figure 3.4: Sampling error in Experiment 1. For illustration purposes, sampling error is mean-averaged across trials for each participant. Marker indicates grand mean and error bars are 95% confidence interval.

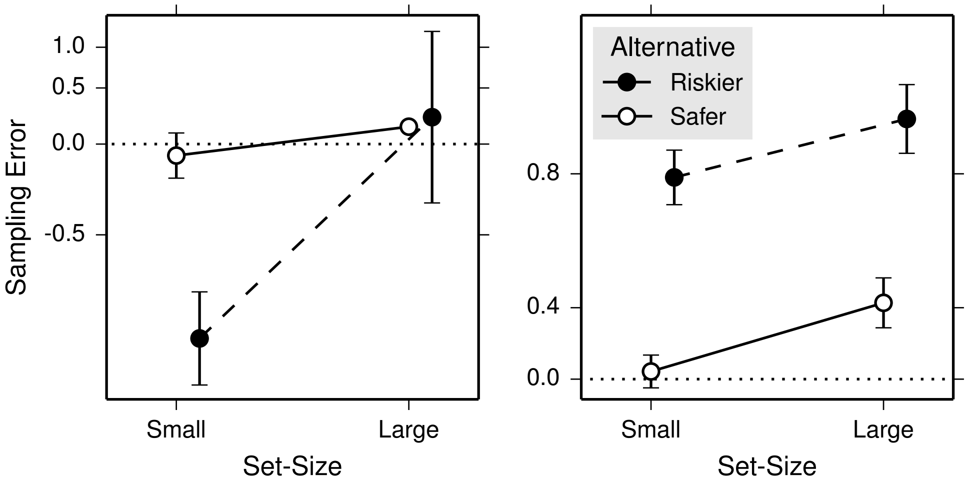
As Experiments 1 and 2 yielded nearly identical results on risk-taking, in what follows we partially pooled the data in the two experiments, allowing effect sizes to differ between the experiments. Only overall effects are reported. To this end, we included by-experiment intercepts and slopes, as well as by-participant intercepts and slopes in all the mixed-effect models reported below.

3.3.2 Sampling error

Sampling error and set-sizes

As discussed above, risk-taking in the sampling paradigm is often attributed to sampling error. Here we examine how sampling error explains the amplified risk-taking. Formally, sampling error can be defined as the difference between a mean sample and an expected pay-off. For example, if an alternative with a .16 probability of a \$5.19 pay-off delivers a series of samples, 0.00, 0.00, and 5.19, the mean sample is $(0.00 + 0.00 + 5.19)/3 = 1.73$. Since this alternative has the expected pay-off of $.16 \times 5.19 = 0.8304$, the sampling error would be $1.73 - 0.8304 = 0.8996$. Positive error indicates that an alternative has produced samples with a higher mean than its expected pay-off; a negative error indicates the opposite.

Sampling error was calculated for each alternative, and we identified the largest sampling error separately for riskier and safer alternatives in each trial and for each participant. Distribution of sampling errors is positively skewed in the



(a) Gain Domain (b) Loss Domain

Figure 3.5: Sampling error in Experiment 2. For illustration purposes, sampling error is mean-averaged across trials for each participant. Marker indicates grand mean and error bars are 95% confidence interval.

gain domain and negative skewed in the loss domain. Therefore in fitting statistical models, we rescaled the sampling error by log-transforming their values after correcting for the different lower and upper bounds for the gain and loss domains, respectively.[1].

Sampling error is displayed in Figures 3.4 and 3.5. These figures illustrate that sampling error increases with a growth in set-sizes, and the increment is largest for riskier alternatives in the gain domain. Mixed-effect linear regressions reveal a significant interaction effect, indicating that the difference between riskier and safer alternatives depends on the interaction between the set-size and the domain: $\chi^2(1) = 51.09$, $\beta = 1.37$, $p < .001$. For the gain domain, the set-size has different effects on riskier and safer alternatives: $\chi^2(1) = 39.10$, $\beta = -1.21$, $p < .001$. While sampling error increases for both riskier and safer alternatives, the increment is larger for a riskier alternative ($\chi^2(1) = 7.85$, $\beta = 1.35$, $p < .01$) than for a safer alternative ($\chi^2(1) = 4.22$, $\beta = 0.13$, $p = .04$).

For the loss domain, the set-size does not have significantly different effects on riskier and safer alternatives $\chi^2(1) = 2.13$, $\beta = 0.16$, $p = .14$. Here, sampling error for riskier alternatives does not differ significantly from sampling error for safer

[1]For the gain domain, we subtracted the smallest possible sampling error, -1.1, and log-transformed the resulting value. For the loss domain where the largest possible sampling error is 1.1, we subtracted 1.1 from the sampling error, and negated to make the distribution positively skewed, log-transformed and negated again. Figures show the data in relation to their untransformed scaling.

alternatives ($\chi^2(1) = 3.48$, $\beta = -0.81$, $p = .06$), but sampling error is significantly larger in a large set than in a small set for both riskier and safer alternatives ($\chi^2(1) = 6.12$, $\beta = 0.47$, $p = .01$).

These results mirror the above results on risk-taking. With a growth in set-sizes in the gain domain, sampling error most drastically increases for a riskier alternative, whose choice is favored by a growth in set-sizes. Indeed, sampling error for a riskier alternative explains choice for riskier alternatives. Specifically, sampling error explains choice differently between the gain and loss domains ($\chi^2(1) = 11.34$, $\beta = -0.83$, $p < .001$) but not between the set-sizes ($\chi^2(1) = 0.47$, $\beta = -0.01$, $p = .49$). Larger sampling error predicts a choice for a riskier alternative in the gain domain ($\chi^2(1) = 21.05$, $\beta = 5.13$, $p < .001$) but not in the loss domain ($\chi^2(1) = 0.11$, $\beta = 0.34$, $p = .11$).

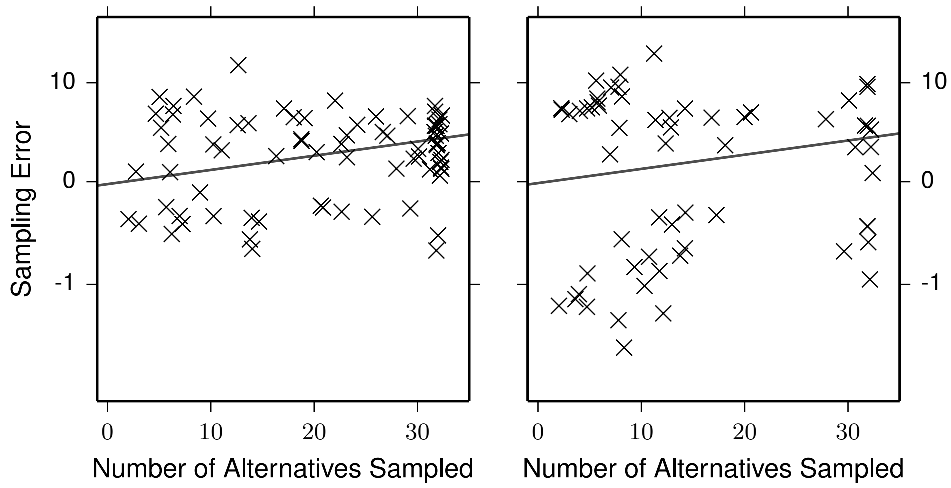
Thus, sampling error significantly increases with a growth in set-sizes in both the gain and loss domains. In the gain domain, however, the increment is more drastic for a riskier alternative. This increment in sampling error for riskier alternatives is linked with more frequent choices for riskier alternatives in a large set.

Sampling error and the number of samples

Sampling error of a riskier alternative is typically associated with the number of samples when set sizes are small (e.g., Hertwig & Pleskac, 2010; Fox & Hadar, 2006; Ungemach et al., 2009). When a riskier alternative is sampled only a few times, the alternative may deliver only a series of 0.00 samples. As a result, the mean sample from a riskier alternative is often lower than its expected pay-off. To investigate relationships between the number of samples and amplified risk-taking, we first test whether the number of samples explains sampling error. As our focus here is on understanding increased sampling error for a riskier alternative and amplified risk-taking in the gain domain, only a riskier alternative in the gain domain is examined below.

Participants sampled 2.16 times from each alternative on average (95% CI [1.24, 3.69])[2], and we entered the number of samples into a mixed-effect linear regression to predict sampling error. Model fit indicates that the effect does not significantly differ between the set-sizes ($\chi^2(1) = 1.84$, $\beta = -0.00$, $p = .17$), and that the number of samples is a non-significant predictor of sampling error ($\chi^2(1) = 3.69$,

[2]This estimate is derived with a mixed-effect Poisson regression, which allows each participant and each experiment to have differing intercepts. As a Poisson regression uses the natural logarithm as a link function, the estimated average is a geometric mean, rather than an arithmetic mean.



(a) Experiment 1 (b) Experiment 2

Figure 3.6: Sampling error for a riskier alternative against the number of alternatives sampled. Each marker represents a trial, and each participant has up to three entries into the figure. Solid line is predicted sampling error from a mixed-effect linear regression.

$\beta = 0.09$, $p = .05$).

Sampling error and the number of alternatives sampled

As a growth in set-sizes is linked with an increment in sampling error, the number of alternatives sampled may also be linked with sampling error in a large set. If individuals sample only two alternatives in a large set, for example, these individuals are functionally restricting themselves to a small set of two alternatives. Thus, the sampling error these individuals encounter should be smaller than the sampling error which other individuals encounter while sampling all 32 alternatives.

Sampling error for a riskier alternative is plotted against the number of alternatives sampled in Figure 3.6. As expected, mixed-effect linear regressions suggest that the number of alternatives sampled explains sampling error for riskier alternatives: $\chi^2(1) = 7.38$, $\beta = 0.03$, $p < .01$. Thus as more alternatives are sampled, sampling error for a riskier alternative increases.

3.4 Discussion of experimental results

As hypothesized, a growth in set-sizes amplifies risk-taking in the sampling paradigm in the gain domain, which in turn eliminates the description-experience gap. Further investigation revealed that this amplification is due to increased sampling error for riskier alternatives.

Sampling error for riskier alternatives is explained by the number of alternatives sampled. As individuals sample more alternatives, at least one alternative tends to deliver a large sample at higher frequency than its underlying probability, resulting in increased sampling error. In the gain domain in particular, sampling error for a riskier alternative increases to a greater extent with a growth in set-sizes than sampling error for safer alternatives, which explains the amplified risk-taking.

Previously, sampling error has been attributed to the number of samples (e.g., Hertwig & Pleskac, 2010; Fox & Hadar, 2006; Ungemach et al., 2009). When a riskier alternative in a small set is sampled only few times, the alternative tends to deliver only 0.00 samples. When a riskier alternative is sampled multiple times, however, the alternative should deliver non-zero samples, increasing the sampling error. This claim is not supported in our results, perhaps because participants in our experiments sampled too few times to observe non-zero samples. Participants in Hertwig et al. (2004)'s study, for example, sampled a median of 7.5 times from each alternative, while participants in our experiments sampled a median of only 2.5 times in Experiment 1 and 1.0 time in Experiment 2.

Thus, our results do not clarify whether sampling error can be alleviated as individuals sample more per alternative. Further, sampling error may also be impacted by the interaction between the number of alternatives sampled and the number of samples. To further explore these effects of sampling on sampling error, the following section describes a simulation using a broader range of samples over a finer range of set-sizes.

3.5 Simulation

To further understand how sampling error impacts on choice probability for a riskier alternative, we examined the influence of set size using three of the best performing models for explaining choices in the sampling paradigm reported in Hau, Pleskac, Kiefer, and Hertwig (2008).

3.5.1 Method

We simulated choices in the environments which reflect those in Experiments 1 and 2, with the following exception: the expected pay-offs for the riskier and safer alternatives were made equal within each trial. This eliminates any confound associated with the relationship between the expected pay-off and the alternative categories (i.e., riskier and safer). Simulated participants followed one of the three models in making a decision: the maximax model, the two-stage model of cumulative prospect theory (the two-stage model henceforth), or the natural mean model. Participants with the maximax model chose the alternative with the largest experienced pay-off.

Participants who followed the two-stage model first weighted the frequency of pay-offs and transformed pay-off amounts into a subjective value to derive the perceived expectation of utility, E:

$$E = w(\text{experienced frequency of pay-off}) \times v(\text{amount of pay-off}),$$

where

$$w(f) = \frac{f^\gamma}{(f^\gamma + (1-f)^\gamma)^{1/\gamma}}$$

and

$$v(x) = \operatorname{sign}(x)\,|x|^\alpha.$$

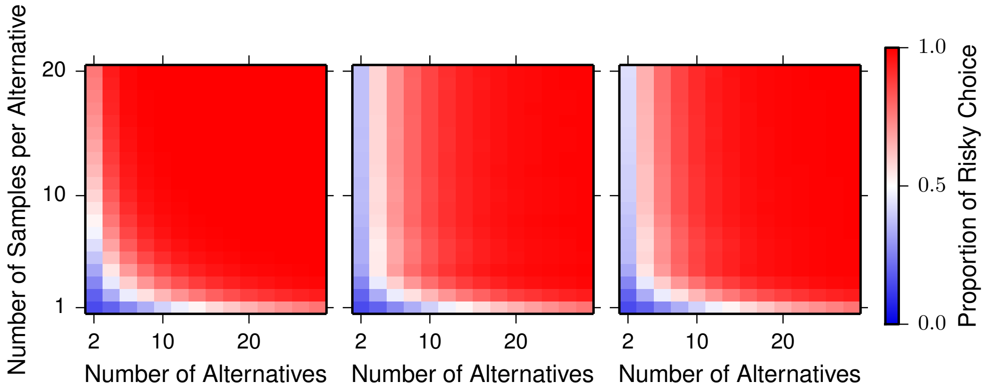
The participants then chose the alternative with the largest perceived expectation of utility. We used parameter values reported in Hau et al. (2008): $\alpha = 0.94$ and $\gamma = 0.99$ for the gain domain; $\alpha = 0.86$ and $\gamma = 0.93$ for the loss domain.

Lastly, participants using the natural mean chose the alternative with the highest mean of sample pay-offs. This natural mean model is a special case of the two-stage model, where $\alpha = 1$ and $\gamma = 1$.

For each of the choice models and for a variety of set-sizes (from 2 to 32) we simulated 10^4 participant choices with a fixed number of samples per alternative.

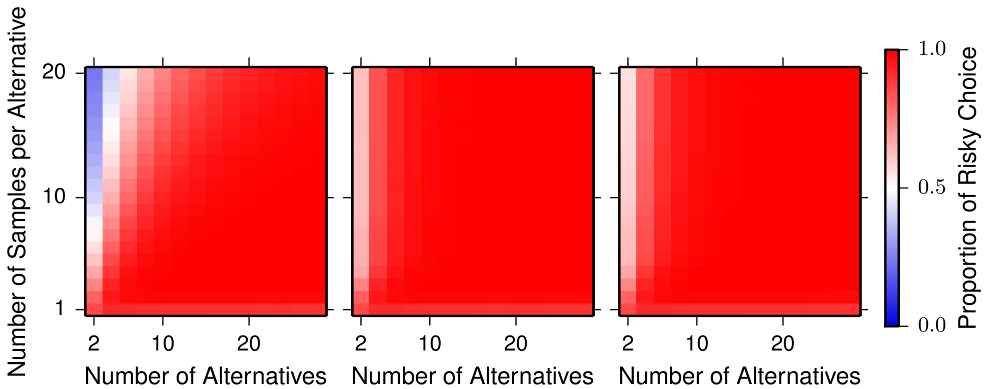
3.5.2 Results and Discussion

The simulation results are summarized in Figures 3.7 and 3.8 for the gain and loss domains, respectively. In both domains the three models make similar predictions over a large area of the simulation space. The most striking result is that, in the gain domain, the simulated participants chose a safer alternative only when the set size and the number of samples are relatively small. When the set-size is larger than two, risky alternatives were chosen almost exclusively. The only exception is when the number of samples is near one per alternative and only when the number of alternatives is equal to or smaller than 12. When more than 12 alternatives are available, all three models predict that a riskier alternative is more frequently chosen, even when the number of samples is as large as 20 samples per alternative.



(a) Maximax (b) Two-stage Model (c) Natural Mean

Figure 3.7: Simulation results for the gain domain.



(a) Maximax (b) Two-stage Model (c) Natural Mean

Figure 3.8: Simulation results for the loss domain.

These results are consistent with the results from Experiments 1 and 2. As sampling error for a riskier alternative increases to a greater extent with a growth in set-sizes than sampling error for a safer alternative, predicted probability of choosing a riskier alternative increases with a growth in set-sizes.

3.6 General Discussion

Psychological experiments often simplify the complexity of the environments in which we live. This simplification is necessary for researchers to isolate and manipulate variables of interest while holding other variables constant. However, this simplification can lead to the neglect of variables most likely to influence behavior outside the laboratory. In the present study, we demonstrate that one such variable — set-size — has a substantial and potentially unavoidable impact on decision making.

Specifically, our experimental results show that a growth in set-sizes facilitates choices for riskier alternatives in the gain domain. This amplified risk-taking eliminates the description-experience gap. Given the number of articles devoted to studying the gap in two-alternative environments, and given the number of real-world environments that are likely to involve more than two-alternatives, the present findings suggest a profound caveat on inferences made from two-alternative environments.

Our results further suggest that the impact of set-sizes on risk-taking is largely due to a change in sampling error. Our findings confirm that sampling error favors safer alternatives when set-sizes are small, but also show that sampling error favors riskier alternatives when set-sizes are large. With a growth in set-sizes, riskier alternatives start appearing better than their expected pay-offs.

Importantly, the simulation indicate that the choice models derived from two-alternative environments predict the amplification of risk-taking associated with a growth in set-sizes. These model predictions are consistent with the observation that subtle differences between choice models may be largely irrelevant in the sampling paradigm, where the statistical structure of information acquisition can have extremely large effects (e.g., Hau et al., 2008). More positively, however, our results demonstrate that the choice models are capable of predicting the observed impact of set-sizes.

The impact of set-size and sampling error is likely to extend beyond choice between monetary pay-offs. The world outside the laboratory is replete with opportunities to make a choice in environments with more than two alternatives. The world wide web now provides consumer ratings and personal experiences for countless potential alternatives, ranging from treatments for depression to suggestions about how to get your partner back. Each of these, in effect, represents a sample, which can provide information about the utility of that choice. Unfortunately, as the number of alternatives increases, the possibility that some alternatives will deliver rare 'lucky' successes increases, appearing as if the utility is unstable across set-sizes.

In the introduction we discussed a specific example with dieting: an individual may assess benefits of diets by recalling other individuals who have tried that diet. Here, recalled individuals are analogous to sampled pay-offs, and the number of diets is analogous to set-sizes. Thus, our findings suggest that the inference on a diet can depend on the number of alternative diets being assessed.

If benefits of diets are rare, assessment on only a few diets may lead to the inference that dieting is a waste of time. Assessment on a larger number of diets, however, is likely to lead to the inference that at least one of the diets is beneficial. Importantly, with more diets considered, it becomes more likely that one of the diets appears much more beneficial than it actually is. Our findings suggest that this effect of sampling error cannot easily be reduced by recalling more individuals.

However, the effect of sampling error can be attenuated by providing feedback to individuals Ert and Erev (2007). If individuals learn the benefit of diet immediately after deciding to try the diet, these individuals may realize a difference between what they expected and the benefits they actually enjoy. If sampling error is large, this difference is also likely to be large, and individuals may eventually learn that the samples they considered were not representative.

Unfortunately, in many environments feedback is not always immediately available. In these cases, the impact of observing rare success can be large and potentially dangerous. Thus, the effects of set-size are extremely important in understanding how individuals make choices in complex many-alternative environments and thus further explored in the next chapter.

Chapter 4

Encouraging myopic choices with information

4.1 Background

When making a choice, people often face a large set of alternatives to choose from. For example, an on-line retailer, amazon.com, now offers more than 3,000 jams. While it may be more likely for people to find a preferred jam in a large set than in a small set, some previous studies report that with more alternatives people are more likely to defer a choice. Broadly, this effect is known as the *too-much-choice* effect. Unfortunately, the prevalence of the too-much-choice effect is disputed as are the factors influencing its occurrence (e.g., Scheibehenne, Greifeneder, & Todd, 2010).

In this study, we investigate how factors which possibly explain choice deferral impact on information search and choice strategies. Previous studies show that information search and choice strategies are often sensitive to choice environments (see Payne, Bettman, & Johnson, 1993, for review). In particular, we examine effects of set size and task complexity, employing risky choice environments. The benefit of employing these environments is that they provides objective measures for determining what information people use to make choices, as well as providing well-defined paradigms for information presentation. Before we describe our approach in more detail, we briefly review the literature on too-much-choice and choice behavior in risky choice environments.

4.1.1 Too much choice effect

The too-much-choice effect is typically associated with three types of behaviors: the first of these is reduced choice quality. When faced with a large set, people often show a reduced capacity to choose the most preferred alternative (Malhotra, 1982; Scammon, 1977). For example, Malhotra (1982) reported that people are less likely to choose the product best aligned with their individual preferences when making a choice between 25 products than when making a choice between 5 products. Second, perhaps as a consequence of this reduced quality choice, people tend to show decreased satisfaction with their choices with a growth in set size (Iyengar & Lepper, 2000). Third, and most relevant to the present study, people are more likely to defer a choice when confronted with a large set (Iyengar, Huberman, & Jiang, 2004).

These behaviors have been attributed to cognitive limitations. Specifically, the information overload hypothesis states that abundant information overloads cognition, and subsequently people become unable to make an informed choice (see Eppler & Mengis, 2004, for review). Previous studies have argued that information overload is induced by large set size, explaining reduced choice quality (Jacoby, Speller, & Kohn, 1974).

A number of studies, however, have shown that large set size alone does not induce information overload. Wilkie (1974), for example, criticizes Jacoby et al. (1974) for not considering choice quality in relation to quality achieved via random choice. When choice quality is compared with quality of random choices, choice quality actually improves with a growth in set size (see also Summers, 1974; Russo, 1974).

Perhaps less controversially, a recent meta-analysis of more than 60 experiments on the too-much-choice effect found that set size have a "virtually zero" effect size on decreased satisfaction and choice deferral (Scheibehenne et al., 2010). Scheibehenne et al. (2010) suggest that in some cases too-much-choice was potentially reliable, but that "to understand the effect that assortment size can have on choice, it will be essential to consider the interaction between the broader context of the structure of assortments — beyond the mere number of options available — and the decision processes that people adopt" (p. 421). Thus, the previously observed influences on choice deferral may be a result of numerous factors and may not necessarily be caused by a growth in set size.

One factor that has since been investigated is the number of attributes (or features) associated with each alternative. Greifeneder, Scheibehenne, and Kleber (2010) demonstrate that a growth in set size alone does not decrease satisfaction, but that increasing the number of attributes along which an alternative is described does. The number of attributes is also indicated to affect choice quality (Helgeson & Ursic, 1993) and choice deferral (Dhar, 1997). (Dhar, 1997), for example, presented participants with two consumer products (e.g., jams) at one time and asked them to make a choice. The results suggest that people are more likely to defer a choice when two products differ in four attributes rather than two attributes.

4.1.2 Risky choice

Studies with risky choice environments ask participants to make choices between alternatives that deliver monetary pay-offs with a fixed probability. For instance, an alternative can be associated with a 90% probability of £3 and a 10% probability of £0. This risky choice has been studied with two separate, yet related, paradigms: the description paradigm and the sampling paradigm. In the description paradigm, pay-off and probability information are explicitly provided at one point in time. For instance, people are told that Alternative A delivers £3 with a 90% probability and £0 with a 10% probability, and that Alternative B delivers £15 with a 20% probability and £0 with a 80% probability.

In contrast, in the sampling paradigm, people typically learn about payoffs by sampling payoffs one-at-a-time over a series of samples. A sequence of samples from Alternative B, for example, could reveal £0, £0, £0, £15, and £0. By observing this sequence, people may learn that Alternative B delivers £15 with a 20% probability. In this sampling paradigm, people can sample each alternative as much as they like before making a choice. Thus, in both description and sampling paradigms, people make choices between alternatives with probabilistic pay-offs, and choices are influenced by the information people gather. With a growth in set size, people must decide how many alternatives to sample and — especially in the sampling paradigm — must decide how many times to sample each alternative.

A number of studies report that choices made in the description paradigm are often distinctively different from those made in the sampling paradigm, especially with respect to risk (e.g., Hertwig et al., 2004; Hertwig & Erev, 2009). In particular, when set size is small, people tend to overweight rare pay-offs in the description paradigm, but people tend to underweight the same rare pay-offs in the sampling paradigm. Though several different explanations have been proposed to explain this difference (e.g., Hertwig & Erev, 2009; Hills & Hertwig, 2010), the differences in choices are largely (though not fully) explained by the number of samples people draw in the sampling paradigm prior to making a choice (Ungemach et al., 2009; Fox & Hadar, 2006): people tend not to sample enough to realize the actual probabilities of pay-off. To realize the 20% probability of £15 pay-off from Alternative B, for example, people have to sample at least 5 times and observe one £15 sample. With other combinations of sampling or pay-offs, people may under- or over-estimate the probability of the £15 pay-off.

In the risky choice environments, information overload can potentially be induced with an increment in the number of branches. One branch corresponds to one possible pay-off and its associated probability. In the above example, both Alternatives A and B have two possible pay-offs (i.e., £3 and £0; and £15 and £0, respectively) and hence have two branches.

The number of branches has been reported to influence a choice in the description paradigm: when set size is large, an alternative with fewer branches is more often chosen, compared to when set size is small (Iyengar & Kamenica, 2010). Iyengar and Kamenica (2010) attribute the avoidance of alternatives with many branches to people's tendency to avoid uncertainty. With a growth in set size, it becomes cognitively taxing to evaluate alternatives with many branches. Thus when faced with a choice between alternatives with many branches, people may avoid evaluating alternatives altogether and simply defer a choice.

In contrast, in the sampling paradigm, the influence of branch has not been as extensively studied. Here, the number of branches in an alternative is only revealed by sampling. As people often do not draw enough samples to realize exact probabilities of pay-off in two-branch alternatives (Ungemach et al., 2009; Fox & Hadar, 2006), people may also be unlikely to draw sufficient samples to recognize the number of branches in an alternative which people sample. Furthermore, people may draw fewer samples per alternative with a growth in set size. Previous studies report that the number of samples per alternative tends to decrease with a growth in set size. According to Hills et al. (2013), for example, when 32 gambles are available, people typically draw only one or two samples per alternative before making a choice. Although research has suggested that people draw more samples from alternative with higher complexity in pay-offs (Lajarraga et al., 2012), we do not expect the number of samples to scale with the number of branches to realize that an alternative has multiple branches or the branch's associated probabilities. Therefore, we predict that an increment in the number of branches is more like to result in frequent choice deferral in the description paradigm than in the sampling paradigm.

In addition in the sampling paradigm, a growth in set size has been reported to increase people's likelihood of sampling rare pay-offs. Chapter 3, for example, report that when confronted with a large set, people are more likely to sample rare but extremely good pay-offs from one of the alternatives, because people tend to sample more alternatives in a large set than in a small set. This experience leads people to evaluate the alternative in a high regard and consequently to choose the alternative. However, in previous studies people do not have the opportunity to defer a choice, and thus their forced choice may not indicate an increased preference for alternatives which delivered these rare pay-offs. In other words, with the opportunity to defer a choice, people may choose deferring a choice over purchasing an alternative that delivers a rare pay-off when that rare pay-off is observed in a large set. However, if previous work (Hills et al., 2013; Chapter 3) does indicate an increased preference, then sampling a rare pay-off in a large set should result in a higher willingness to purchase the alternative presenting the rare pay-off. We hypothesize that a growth in set size will result in more frequent choice purchase, as opposed to choice deferral, in the sampling paradigm.

4.1.3 Information search and choice strategy

As stated above, the utility of the description and sampling paradigms is that they allow us to quantitatively examine the "broader context of the structure of assortments" and the "decisions processes people adopt" in environments that may induce too-much-choice, as called for by Scheibehenne et al. (2010). We discussed one aspect of structure above — the number of attributes per alternative. Choices in the description and sampling paradigms, however, also involve other two components: information search and information processing.

As information search has direct influence on the quality of the information available for processing, we predict that information search will play a critical role in choice (see e.g., Hills & Hertwig, 2010). As presented information differs between the description and sampling paradigms, we further predict that information search will have differential effects on choice in the two paradigms. Specifically, we hypothesize that more ready reliance on *thin-search* — spreading out and thinning information search across available alternatives — will impact on a choice in two distinct ways in the two paradigms.

In the description paradigm, reliance on thin-search leads to incomplete information about alternatives and therefore results in use of non-compensatory choice strategies: people search and use only a a subset of information on probabilities and pay-offs per alternative. Previous work reports that when consumer products are described in terms of many attributes, people search only a subset of attributes (Cook, 1993) and choice strategies tend to be non-compensatory (Ford, Schmitt, Schechtman, Hults, & Doherty, 1989). Thus, we predict that the consequence of relying on thin-search — use of non-compensatory strategies — will be most pronounced as the number of branches per alternative increases.

To investigate strategy use in the description paradigm, we follow previous research on strategy selection in risky choice. In particular, (Thorngate, 1980) and (Payne et al., 1993) provide us with a systematic way to explore various strategies. Among them, we examine three strategies: expected pay-off, maximax and lexicographic. According to the expected pay-off strategy, people choose an alternative with the largest expected pay-off. Thus, the expected pay-off strategy is compensatory: it requires all the information (i.e., probability and pay-off in all branches) to be processed to make a choice. Although other strategies have been proposed (e.g., see Hau et al., 2008), these three strategies are selected from the best models in explaining choices (Hau et al., 2008).

In contrast, the maximax and lexicographic strategies are non-compensatory: the maximax strategy predicts that people choose an alternative whose best pay-off is largest within a set. Thus the maximax strategy requires consideration of only pay-offs and not probabilities of pay-offs. If people employ the lexicographic strategy, they choose an alternative whose most likely pay-off is largest. Thus, the lexicographic strategy requires consideration of probabilities of pay-offs and only one possible pay-off per alternative. These two strategies do not require all the information to be processed and hence are non-compensatory.

In the present context, these strategies determine likelihood that an alternative will exceed a given threshold for purchase. For example, with the expected pay-off strategy, for example, a choice is deferred when all the alternatives have an expected pay-off below a threshold for purchase. With the maximax strategy, a choice is deferred when all the alternatives in a set have maximum pay-off of below a threshold.

In contrast to the description paradigm, strategy use in the sampling paradigm is less important with reliance on the thin-search. This is because fewer samples provide less information about probabilities; in the limit, all three choice strategies listed above make the same prediction if people take only one sample per alternative. Instead, the increased likelihood of sampling a rare pay-off (Chapter 3) will likely have a larger impact on a choice; the question remains, however, as to whether or not an alternative with a rare pay-off is preferred over deferring a choice altogether.

4.1.4 Present study

In the following experiment, we tested the following hypotheses: 1) we expect choice deferral to be more frequent when an alternative has many branches than when an alternative has few branches, only in the description paradigm; 2) an increment in the number of branches leads to thin-search and consequently use of a non-compensatory strategy in the description paradigm; 3) a growth in set-sizes results in thin-search in the sampling paradigm; and 4) thin-search in the sampling paradigm is associated with less frequent choice deferral. In addition, we explored relationships between strategy use and choice deferral in the description paradigm.

4.2 Method

4.2.1 Participants

Seventy-two (63 females, 8 males, and 1 undisclosed) undergraduate students participated in the study. Prior to collecting the data, we decided to recruit all the 88 students who are offered the chance to participate in an experiment for a course credit. Out of 88, 16 students did not participate. Their age ranges from 18 to 33, with a mean of 19.1.

4.2.2 Design

The experiment employed a 2 (between-participants, set-size: small or large) $\times$ 2 (between-participants, paradigm: description or sampling) $\times$ 2 (within-participant, branch: few or many) design. Participants were randomly assigned to each of between-participant conditions.

4.2.3 Apparatus

Alternatives were independently and randomly generated for each trial for each participant with the following procedure. First, an expected pay-off is determined with a random draw from the normal distribution whose mean is 1.00 and standard deviation is 0.30. This normal distribution is truncated to ensure that no alternative is assigned to a negative expected pay-off. The expected pay-off is randomly divided into two (two branch condition) or four (four branch condition), with a constraint that one of the branches is assigned to 0.00.

This constraint makes it unfeasible to use other possible strategies than the expected pay-off, maximax, and lexicographic strategies. Minimax strategy, for example, predicts that an alternative with the highest minimum pay-off is chosen. With one branch always assigned to the pay-off of 0.00, the minimum pay-off is 0.00 for all alternatives, making the minimax strategy unable to discriminate alternatives.

For each alternative, an expected pay-off was divided into branches by multiplying with multinomial probabilities randomly drawn from Dirichlet distribution whose number of concentration parameter is one fewer than the number of branches and whose concentration parameter is all 1.00. Then, the probability of pay-off is independently determined with a random draw from Dirichlet distribution whose number of concentration parameters equals to the number of branches and whose concentration parameters is all 1.00. Then, the randomly divided numbers were further divided by the probability of pay-off to derive the pay-off amount. The probability and the amount of pay-off are both rounded to the nearest two decimals.

In generating an alternative with four branches, for example, a random draw from the normal distribution can be 1.20. This 1.20 is treated as an expected pay-off from this alternative and randomly divided into four numbers: for instance, 0.62, 0.41, 0.17, and 0.00. Independently drawn random numbers from Dirichlet distribution may be .34, .13, .08, and .45. Then, this alternative is associated with a .34 probability of 1.82 (= 0.62/0.34), a .13 probability of 3.15 (= 0.41/0.13), a .08 probability of 2.13 (= 0.17/0.08), and a .45 probability of 0.00. Another alternative for the same trial is independently generated with the same procedure.

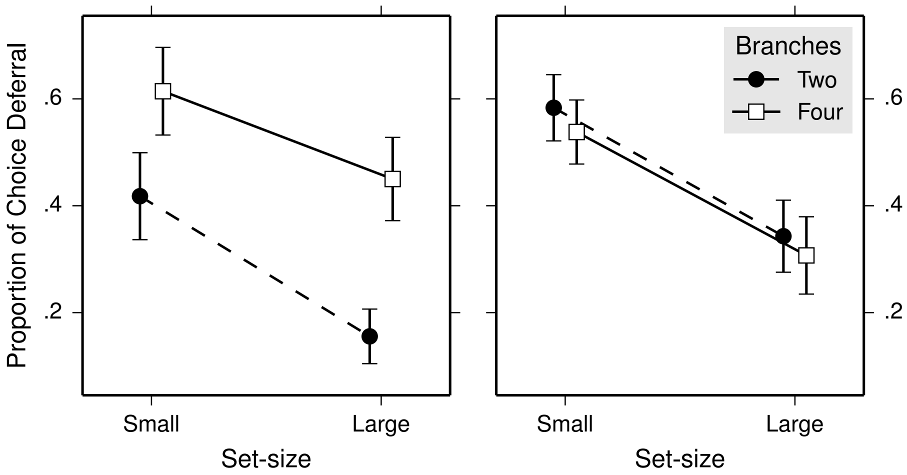
4.2.4 Procedure

Participants were instructed that their bonus payments would depend on their choices during the experiment. The experiment asked participants to make 10 choices in total, 5 of which were choice between alternatives with two branches and the other 5 were with four branches.

Each trial displayed 2 alternatives (small set-size) or 35 alternatives (large set-size) as an array of boxes on a screen. Participants were asked to sample from the alternatives as many times as they wanted and then decide whether to choose an alternative to purchase with £1.00 or to defer a choice and keep £1.00.

Every time an alternative was sampled, information about the alternative was presented for 1,000ms. In the description paradigm, the information displayed the probability and pay-off amount (e.g., 20%, £15.00; 80%, £0.00). In the sampling paradigm, the information presented was a random draw with replacement from the pay-off distribution associated with that alternative. For example, when an alternative with a 20% probability of £15.00 was sampled in the sampling paradigm, 2 in 10 on average displayed £15.00, otherwise £0.00. After sampling if participants decide to purchase an alternative, participants were then asked to indicate the alternative whose draw they wish to purchase.

Participants did not learn about the pay-offs from their purchases until the end of the experiment, when one of the 10 trials was randomly selected. If participant purchased a draw in the randomly selected trial, participant was paid the



(a) Description Paradigm (b) Experience Paradigm

Figure 4.1: Proportion of choice deferral. Error bar represents standard error.

pay-off they earned from the purchase. If participant did not purchase a draw in the randomly selected trial, participant was paid participation fee of £1.00. The payment ranged from £0.00 to £2.39, and its mean was £0.61.

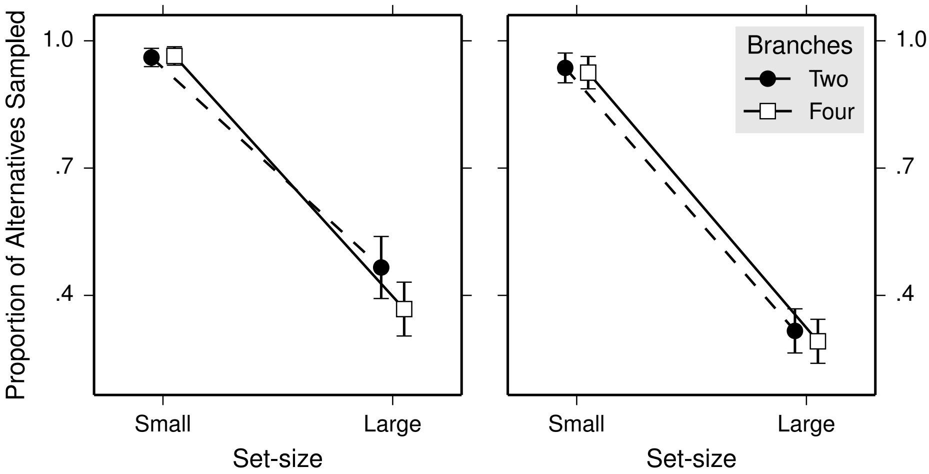
4.3 Results

4.3.1 Choice deferral

Figure 4.1 presents the proportion of deferred choices as a function of paradigm (description or sampling), set-size (small or large) and branch (few or many). Figure 4.1 shows that the proportion of choice deferral fell with a growth in set-sizes in both paradigms. Moreover, consistent with our hypothesis, an increment in the number of branches increased the proportion of choice deferral in the description paradigm. In the sampling paradigm, however, the number of branches does not appear to impact on the proportion of choice deferral.

A mixed-effect logistic regression[1] suggests a significant effect of set-size on choice deferral: $\chi^2(1) = 6.13$, $\beta = -0.51$ (95% CI [-0.92, -0.11]), $p = .01$. In addition, an interaction effect indicates that an effect of number of branches significantly depends on paradigm: $\chi^2(1) = 11.35$, $\beta = -1.32$ (95% CI [-2.07, -0.57]), $p < .001$. In the description paradigm, choice is significantly more likely to be de-

[1]All the mixed-effect regressions we report here include maximal random factors: by-participant intercept and slopes, and, if applicable, by-condition intercept and slopes. We assess statistical significance by examining model fit with chi-square tests.



(a) Description Paradigm (b) Experience Paradigm

Figure 4.2: Proportion of alternatives sampled. Error bar represents standard error.

ferred in the four branch condition than in the two branch condition: $\chi^2(1) = 14.82$, $\beta = 1.33$ (95% CI [0.66, 1.99]), $p < .001$. In the sampling paradigm, branch has a non-significant effect: $\chi^2(1) = 0.14$, $\beta = -0.09$ (95% CI [-0.57, 0.39]), $p = .71$. The other interaction and main effects are non-significant: $ps > .24$.

These results show that a growth in set-sizes reduces choice deferral in both paradigms. Also, consistent with our hypothesis (Hypothesis 1), increasing the number of branches induces choice deferral only in the description paradigm. In the following sections, we investigate factors driving these effects. First, we examine whether a growth in set-sizes induces thin-search.

4.3.2 Thin-search

To investigate thin-search, we calculated the proportion of alternatives sampled and the number of samples per alternative for each trial for each participant. In counting the number of samples, we only considered alternatives which had sampled at least once. The proportion and the counts are summarized in Figures 4.2 and 4.3.

Figure 4.2 shows that the proportion of alternatives sampled drops with a growth in set-sizes. These proportions were multiplied with 0.90 to handle 1s, added 0.05 to handle 0s and then logit-transformed. We examine the transformed proportion with a mixed-effect linear regression. Model fit indicates that proportion is significantly smaller in a large set than in a small set: $\chi^2(1) = 80.53$, $\beta = -3.19$ (95% CI [-3.69, -2.69]), $p < .001$. Interaction effects and other main effects are all non-significant: $ps > .10$.

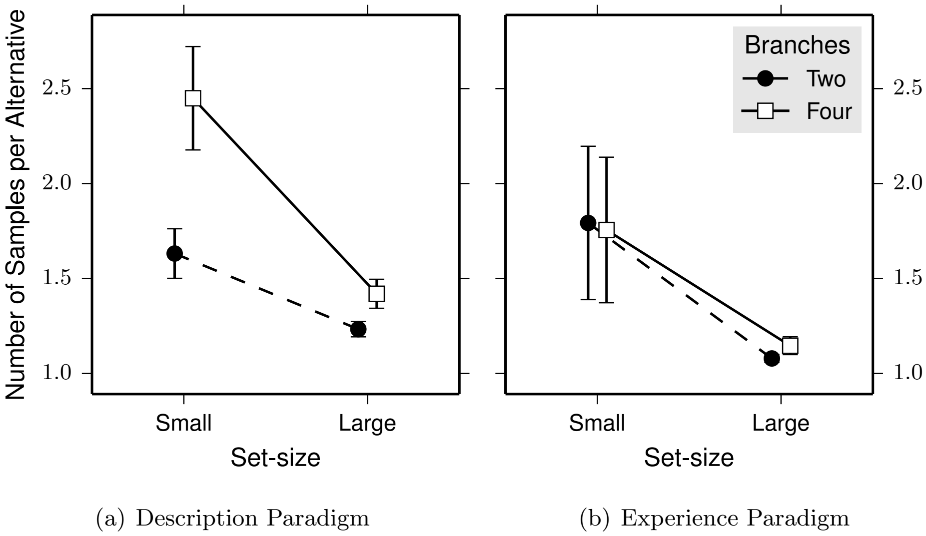


Figure 4.3: Number of samples per alternative. Error bar represents standard error.

Similarly, Figure 4.3 shows that the number of samples drops per alternative as set-sizes grow. In the description paradigm, the number of samples is the number of times participant clicked on an alternative. As one click displayed information on an alternative for 1,000ms, the number of samples in the description paradigm corresponds to the number of seconds participant could spend on examining an alternative.

The number of samples is examined with a mixed-effect Poisson regression, separately for the two paradigms. In the description paradigm, the number of samples differs between both set-size and number of branches, as indicated by the interaction effect: $\chi^2(1) = 4.39$, $\beta = -0.24$ (95% CI [-0.45, -0.02]), $p = .04$. The number of samples is significantly smaller when a set is large than when a set is small: $\chi^2(1) = 9.91$, $\beta = -0.26$ (95% CI [-0.41, -0.10]), $p < .01$ for the two branch condition; and $\chi^2(1) = 16.18$, $\beta = -0.51$ (95% CI [-0.72, -0.30]), $p < .001$ for the four branch condition. Also, the number of samples is significantly larger in the many branch condition than in the two branch condition: $\chi^2(1) = 23.40$, $\beta = 0.42$ (95% CI [0.25, 0.59]), $p < .001$ for a small set; and $\chi^2(1) = 5.30$, $\beta = 0.14$ (95% CI [0.03, 0.24]), $p = .02$ for a large set.

Thus in the description paradigm, more time is spent on evaluating an alternative with four branches than an alternative with two branches. With a growth in set-sizes, however, less time is spent on evaluating an alternative, regardless of the number of branches. These results imply that an individual may be able to process less information per alternative when set-sizes are large. We examine this implication in the next subsection.

In the sampling paradigm, the difference in the numbers of samples between the set-sizes does not significantly differ between branch conditions, as indicated by the non-significant interaction effect: $\chi^2(1) = 0.46$, $\beta = 0.07$ (95% CI [-0.14, 0.29]), $p = .50$. The number of samples is significantly smaller in a large set than in a small set: $\chi^2(1) = 4.79$, $\beta = -0.24$ (95% CI [-0.34, -0.14]), $p < .001$. The number of samples, however, does not differ significantly between the few and four branch conditions: $\chi^2(1) = 0.62$, $\beta = 0.03$ (95% CI [-0.05, 0.12]), $p = .43$. Thus in the sampling paradigm, a choice is based on fewer samples per alternative when set-sizes are large, supporting the hypothesis that thin-search becomes more prevalent with a growth in set-sizes (Hypothesis 3).

The results thus far support thin-search: less information is sampled in both description and sampling paradigms with a growth in set-sizes. In the next analysis, we examine how thin-search relates to choice deferral.

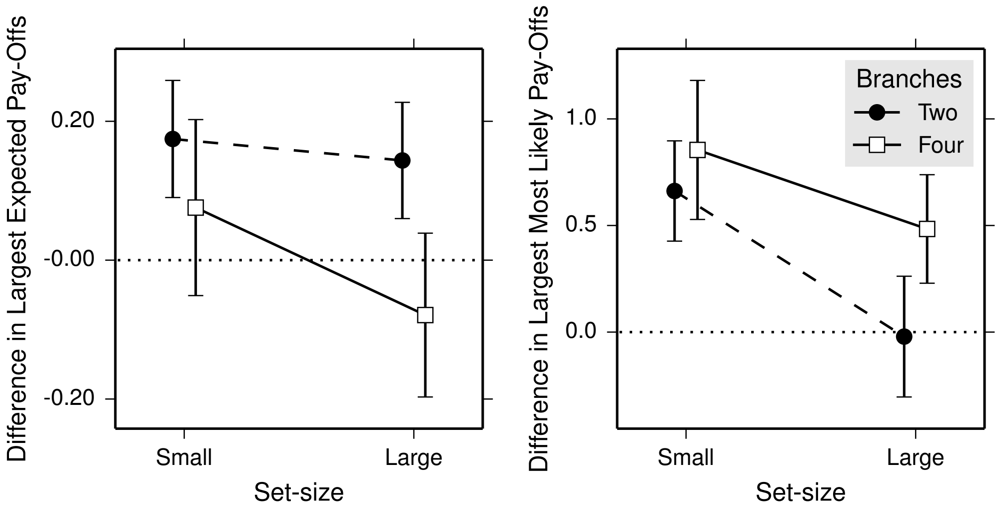
4.3.3 Choice deferral in the description paradigm

In the following analysis, we investigate which strategy an individual uses to decide whether to defer a choice. As discussed above, we examine the expected pay-off, maximax, and lexicographic strategies. According to the expected pay-off strategy, choice deferral is explained by the largest expected pay-off among alternatives sampled. Also according to the maximax and lexicographic strategies, choice deferral is explained by the largest maximum pay-off and the largest most likely pay-off, respectively.

We identified expected pay-off, maximum pay-off, and most likely pay-off for each alternative sampled, and determined the largest expected pay-off, the largest maximum pay-off, and the largest most likely pay-off for each trial for each participant. The maximum pay-off and the most likely pay-off, however, tend to have different amounts between the few and four branch conditions: because the expected pay-off from an alternative is determined independently of the number of branches, pay-off from one branch tends to decrease as the number of branches increases.

Thus to assess the maximax and lexicographic strategy independently of set-sizes, we normalized these pay-offs within each set-size, so that their means are 0.00 and their standard deviations are 1.00. The differences in largest expected pay-offs and largest most likely pay-offs are summarized in Figure 4.4.

The left panel in Figure 4.4 displays difference in largest expected pay-offs when a choice was purchased compared to when it was deferred. Here, a positive value along the vertical axis indicates that the largest expected pay-off within a



(a) the expected pay-off strategy

(b) Lexicographic Strategy after Normalisation

Figure 4.4: Differences in largest pay-offs in the description paradigm when a choice was purchased compared to when a choice was deferred. Pay-off in the right panel is normalized. Error bar represents standard error.

trial is greater prior to choice purchase than prior to choice deferral. Difference of 0 indicates that the largest expected pay-off does not differ between prior to choice purchase and choice deferral, and also, a negative value indicates that the largest expected pay-off is smaller prior to choice purchase than prior to choice deferral. In general, positive difference indicates that the expected pay-off strategy explains choice deferral. Thus, the left panel in Figure 4.4 shows that the expected pay-off strategy explains choice deferral in the two branch condition for both small and large sets.

Similarly, the right panel in Figure 4.4 illustrates difference in largest most likely pay-offs. This panel shows that the lexicographic strategy explains choice deferral, except when an alternative has two branches in a large set.

To statistically confirm these observations, the pay-offs are entered into a mixed-effect logistic regression to predict choice deferral. To explore different strategy use across set-sizes and the number of branches, we include interaction terms with set-size and branch in the regression.

An interaction effect indicates that the expected pay-off strategy explains choice deferral better in the two branch than in the four branch condition: $\chi^2(1) = 11.47$, $\beta = 1.12$ (95% CI [0.47, 1.76]), $p < .001$. The expected pay-off strategy significantly explains choice deferral in the few branch condition ($\chi^2(1) = 4.62$,

$\beta = -1.92$ (95% CI $[-3.76, -0.07]$), $p = .03$) but not in the four branch condition ($\chi^2(1) = 3.57$, $\beta = -3.41$ (95% CI $[-7.50, 0.67]$), $p = .06$). The other interaction effects are non-significant: $ps > .06$.

Also, the model fit indicates that the maximax strategy explains choice deferral: $\chi^2(1) = 4.84$, $\beta = -0.34$, $p = .03$. This effect of maximax strategy does not depend on set size or branch condition: $ps > .07$.

Importantly for the thin-search prediction, the effect of the lexicographic strategy depends on both set size and the number of branches: $\chi^2(1) = 9.67$, $\beta = 1.98$, $p < .01$. The lexicographic strategy explains choice deferral in the four branch condition: $\chi^2(1) = 6.41$, $\beta = -0.61$, $p = .01$. This strategy, however, explains choice deferral better in a small set than in a large set when an alternative has two branches: $\chi^2(1) = 6.28$, $\beta = 1.20$, $p = .03$. While lexicographic strategy does not significantly explain choice deferral in the two branch condition in a large set ($\chi^2(1) = 1.71$, $\beta = -0.39$, $p = .19$), it does in the four branch condition ($\chi^2(1) = 12.87$, $\beta = -1.67$, $p < .01$).

As only the effects of the expected pay-off and lexicographic strategies significantly differ between conditions, these results imply that people switch between the expected pay-off and lexicographic strategies depending on the structure of choice environments.

Thus to further examine the change in strategy use, we classified each participant as a user of the expected pay-off strategy or lexicographic strategy for each branch condition. This classification is based on log-likelihood of the mixed-effect logistic regressions fitted separately for each branch condition and each set size. Specifically, we first predicted the probability of choice deferral using the fitted models, computed the log-likelihood of the participant's response, and compared the log-likelihood with the expected pay-off strategy against the log-likelihood with the lexicographic strategy. If the log-likelihood was larger with the expected pay-off strategy for a participant, indicating that the expected pay-off strategy is a better predictor of choice deferral than the lexicographic strategy, we classified this participant as a user of the expected pay-off strategy. The number of participants classified into each strategy use is displayed in Table 4.1. As branch is a within-participant condition, each participant is classified into one of the four cells in this table.

Table 4.1 shows that the largest proportion (.53) of participants is classified as using the expected pay-off strategy in the two branch condition but as using the lexicographic strategy in the four branch condition. A Poisson regression confirms that significantly more participants are classified as users of the expected pay-off strategy than the lexicographic strategy in the few branch condition: $\beta = 0.79$ (95%

		Four Branch Condition	
		Expected Pay-off	Lexicographic
Two Branch Condition	Expected Pay-off	5 (.16)	17 (.53)
	Lexicographic	4 (.13)	6 (.19)

Table 4.1: Number (and proportion) of participants classified into each strategy use.

CI [0.07, 1.58]), $p = .03$. In the four branch condition, however, significantly more participants are classified as users of the lexicographic strategy than the expected pay-off strategy: $\beta = -2.39$ (95% CI [-1.76, -0.20]), $p = .01$. Other effects, including effects of set-size, are non-significant, $ps > .28$.

This difference in use of the strategies explains the increased proportion of choice deferral with an increment in the number of branches, as we saw in Figure 4.1. Prior to choice purchase, the largest expected pay-off from an alternative with two branches is £1.35 on average (95% CI [1.30, 1.40]) and the largest most likely pay-off from an alternative with many branches is £1.24 on average (95% CI [1.14, 1.35]). The overlapping confidence intervals indicate that a threshold for the expected pay-off strategy is similar to or potentially the same as a threshold for the lexicographic strategy.

Across all the trials, where choice is purchased and deferred, the largest expected pay-off in the two branch condition is significantly greater than the largest most likely pay-off in the many branch condition: $\chi^2(1) = 4.59$, $\beta = -0.13$ (95% CI [-0.26, -0.01]), $p = .03$. Thus if thresholds are similar or the same for the expected pay-off and lexicographic strategies, at least one alternative in a set is more likely to be above a threshold for the expected pay-off strategy in the two branch condition than for the lexicographic strategy in the four branch condition. Consequently, use of the expected pay-off strategy in the two branch condition is more likely to result in choice purchase, hence less likely to result in choice deferral, than use of the lexicographic strategy in the four branch condition.

Also with a growth in set-sizes, the largest expected pay-off increases for an alternative with few branches across all the trials: $\chi^2(1) = 23.58$, $\beta = 0.28$ (95% CI [0.18, 0.37]), $p < .001$. Similarly, the largest most likely pay-off increases for an alternative with many branches with a growth in set-sizes: $\chi^2(1) = 23.58$, $\beta = 0.28$ (95% CI [0.18, 0.37]), $p < .001$. These increments are due to the way alternatives are generated: Variance in pay-offs increases as more alternatives are generated. Thus, effects of set-sizes are due to the property of experiment and particular pay-off to which an individual attends.

Importantly, once the effects of expected pay-off and lexicographic strategies

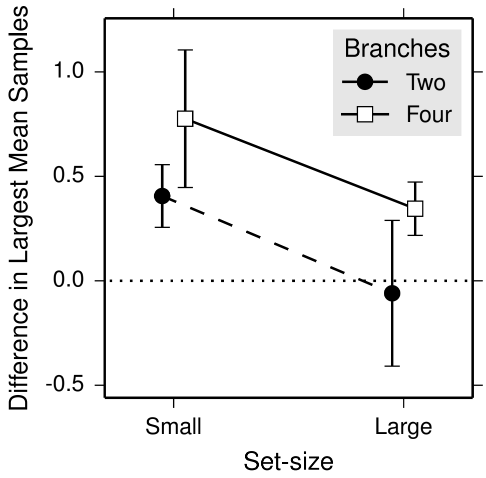


Figure 4.5: Differences in maximum mean samples in the sampling paradigm. Error bar represents standard error.

are controlled for, the effect of set-size and branch on choice deferral becomes non-significant: $\chi^2(1) = 3.05$, $\beta = -1.13$ (95% CI $[-2.41, 0.15]$), $p = .08$; and $\chi^2(1) = 1.49$, $\beta = 1.88$ (95% CI $[-1.12, 4.90]$), $p = .22$, respectively. These results indicate that the higher likelihood to see a larger pay-off in a large set explains the reduced proportion of choice deferral with a growth in set-sizes. Also, an increment in the number of branches promotes use of a non-compensatory strategy use, which results in the increased proportion of choice deferral.

4.3.4 Choice deferral in the sampling paradigm

As in the above analysis, we consider three possible strategies in the sampling paradigm: natural mean, maximax and lexicographic strategies. The natural mean strategy predicts that if the largest mean samples from an alternative is less than a threshold for choice purchase, an individual defers a choice. Similarly, the maximax strategy posits that a largest sample is less than a threshold, choice is deferred. Lastly according to the lexicographic strategy, if the largest most frequent sample is less than a threshold, choice is deferred.

These three strategies make nearly identical predictions, given the relatively small number of samples per alternative (Figure 4.3). Thus, we report only the results with the natural mean strategy. As in the above analysis, we summarize the differences between choice purchase and choice deferral in Figure 4.5. Positive value in this figure indicates that the largest mean sample is greater prior to choice purchase than prior to choice deferral, supporting predictions from the natural mean strategy.

These largest mean samples are examined with a mixed-effect logistic regression, which includes an interaction term with set-size. Model fit indicates that choice deferral is significantly predicted by largest mean samples: $\chi^2(1) = 85.67$, $\beta = -1.44$ (95% CI $[-2.05, -0.81]$), $p < .001$. This effect does not significantly depend on set-size or the number of branches: $ps > .11$. Thus, observing a large sample leads to choice purchase.

Replicating previous research (Noguchi & Hills, under review), the largest mean sample is greater in a large set than in a small set: $\chi^2(1) = 18.97$, $\beta = 1.79$ (95% CI $[1.12, 2.46]$), $p < .001$. This is because an individual is more likely to encounter a large but rare sample as he or she draws samples from more alternatives. As a large sample is more likely to be observed in a large set, the proportion of choice deferral decreases with a growth in set-sizes.

Once the effect of natural mean strategy is controlled for, the effect of set-size on choice deferral becomes non-significant: $\chi^2(1) = 1.40$, $\beta = -0.58$ (95% CI $[-1.54, 0.37]$), $p = .24$. Thus, a growth in set-sizes leads to thin-search in the sampling paradigm, which in turn explains less frequent choice deferral in a large set. The results are consistent with Hypothesis 4.

4.4 Discussion

The too-much-choice effect suggests that a growth in set size reduces choice quality, decreases satisfaction, and increases frequency of choice deferral. These effects were initially attributed to large set size, but empirical evidence suggest that the picture is more subtle. In particular, in the present work we found that choice deferral is mediated by the interaction of information presentation, information complexity, and information search. In what follows, we briefly describe these interactions and their potential implications for future research.

First, our results demonstrate that in both description and sampling paradigms, when confronted with a large set of alternatives, people sample more alternatives, but appear to acquire less information per alternative. Thus, people respond to large sets by using thin-search. This finding is consistent with previous findings on task complexity and contingent processing. When confronted with many apartments to choose from, for example, people tend to quickly eliminate certain apartments from consideration, and more closely examine the remaining apartments (Payne, 1976; Lussier & Olshavsky, 1979). We observed a similar kind of elimination, but in our experiments people eliminate alternatives from consideration without investigating them at all. Such stopping rule of information search in different domains is of broad general interest (e.g., Seale & Rapoport, 1997; Browne & Pitts, 2004) and a clear direction for future research. However, in this work we were specifically interested in the implications of thin-search on choice deferral.

Second, we found that choice deferral is less frequent in a large set for both description and sampling paradigms. In part, this is consistent with prior work suggesting that people are more likely to find an alternative they prefer in a large set than in a small set, if they have sufficiently well-defined preferences (Chernev, 2003; Scheibehenne et al., 2010). We note that this explanation is to some degree unsatisfactory, because if people defer a choice one can always argue that preferences were *not established well enough*. If people do not defer a choice, one can simply argue the opposite. Given that people routinely exhibit choice strategies that sometimes favor likely (lexicographic) and sometimes favor maximum outcomes (maximax), as well as many other variations, it is difficult to conclude (except in retrospect) that people had well-defined preferences before they completed our study. Instead, we focused on environmental factors influencing strategy use, which explain choice deferral.

We find that a non-compensatory strategy is likely to be used in the description paradigm, especially when an alternative is associated with many branches. This non-compensatory strategy use may appear in conflict with previous findings. Glöckner and Herbold (2011), for example, analyzed eye-movement process data and reported support for compensatory strategy use: people consider all the information presented. In Glöckner and Herbold (2011)'s study, however, alternatives had only two branches. Our results indicate that use of a compensatory strategy is most prevalent when the number of branches is few. With an increment in the number of branches, use of a non-compensatory strategy becomes more prevalent. This finding is consistent with prior work (e.g., Ford et al., 1989), where a non-compensatory strategy is often more readily used with an increment in the number of attributes to describe alternatives.

In the sampling paradigm, thin-search increases the likelihood of encountering rare pay-offs, which often results in choice purchase. Similar to choices in the description paradigm, choices based on thin-search can also be considered to follow a non-compensatory strategy. Of course, as has frequently been observed in the sampling paradigm, information search is rarely sufficient to provide people with accurate probability information (Hertwig et al., 2004; Ungemach et al., 2009). One of the principle contributions of the present work is to show that thin-search leads to an increase in preference for alternatives with rare pay-offs when set sizes is large.

With thin-search, people could have felt uncertain about possible pay-offs from alternatives, disliked this uncertainty, and avoided purchasing an alternative (e.g., Iyengar & Kamenica, 2010). Our results, however, show that despite the uncertainty, people often decide to purchase an alternative. Thus, frequent choice for alternatives with rare large pay-offs in a large set, reported in previous work (e.g., Hills et al., 2013; Chapter 3), is not a result of forced choice — people appear to develop preferences for alternatives with rare pay-offs sufficient to purchase an alternative. The present study, however, did not measure choice satisfaction, and it could be possible that choice satisfaction is lower when an alternative with rare large pay-offs is purchased, compared to when an alternative with more frequent but small pay-offs is purchased. Examination of this possibility is left for future research.

Our results, however, do not show psychological factors behind thin-search. When faced with large sets, for example, people may attempt to reduce cognitive effects associated with sampling by reducing the number of samples per alternative. Alternatively, people may feel exhausted after making one choice in a large set, and subsequently when faced with a large set, people may be left with little cognitive resource to evaluate alternatives as rigorously as they could have otherwise.

One place where thin-search appeared to *increase* frequency of choice deferral is in the description paradigm with increased branches per alternative. We suggest this increase is potentially due to people having a similar threshold for choice purchase for both compensatory and non-compensatory strategies. To illustrate, suppose people are choosing whether to purchase an assorted jam box. At a given price, a box can have many small containers of jam or fewer larger containers of jam. When a box contains many jams, people using a non-compensatory strategy may inspect only a few jams and infer that no jam which they prefer is in sufficient quantity to warrant purchase. This limited attention may lead people to regard the box to be overpriced and defer a purchase. In contrast when a box contains few jams, people are more likely to employ a compensatory strategy, observe greater amounts of preferred jam when it is present, and proceed to purchase a box. This example highlights that subjective utility of an alternative may be sensitive to structures of choice environment. This unstable nature of utility is further demonstrated in the following chapters.

The number of attributes, however, is not likely to be the only factor driving non-compensatory strategy use: Previous studies report that the cost of information search is also related to non-compensatory strategy use. Bröder (2000) showed that search costs make it more likely that people will search only a subset of attributes and consequently use a non-compensatory strategy.

To summarize, the present work examined two of the possible factors of choice deferral: set size and the number of branches. Our results show that set size and the number of branches impact on information search and choice strategies, but that the exact impact differs between the description and the sampling paradigms. In both paradigms, set size is related to thin-search — search over more alternatives but for less information per alternative. In the description paradigm, increasing the number of branches is associated with use of a non-compensatory strategy, which often leads to choice deferral. In contrast, thin-search often leads to choice purchase in the sampling paradigm. These findings demonstrate the importance of interactions between the choice environments, information search, and choice strategies, in understanding choice deferral. Thus, the present work provides a potential answer to the concern raised by Scheibehenne et al. (2010) and suggest that choice deferral as observed in previous work (e.g., Iyengar & Lepper, 2000) is a function of interactions between multiple factors.

Chapter 5

In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions

5.1 Background

In the domain of choice between multiple alternatives, the attraction, compromise, and similarity effects demonstrate some puzzling behaviors. Together these effects demonstrate that individuals do not choose by selecting the alternative with the highest value or utility. Instead, individuals choose as if the value or utility of an alternative is temporarily affected by the other alternatives in the choice set they face. This is puzzling because how much an individual enjoys the car she or he buys, for example, should be independent of the cars he or she does not buy. These context effects are often interpreted as indicating that a choice is reached by comparing available alternatives. This study investigated how alternatives are compared, using eye movement data collected while people make a series of three-alternative choices.

To illustrate the attraction, compromise, and similarity effects, suppose an individual is choosing among different cars. Available cars are described in terms of the two attributes, quality and economy, where Car A is better on the quality dimension but Car B is better on the economy dimension (Figure 5.1). The attraction effect is produced by adding Car D to the choice of Cars A and B. Car D is inferior

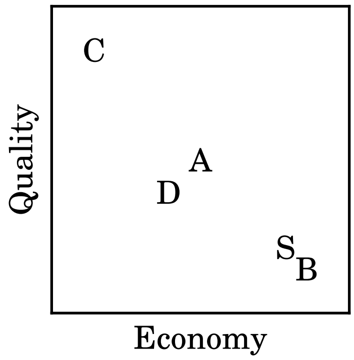


Figure 5.1: Illustration of various alternatives. The probability of A being chosen over B can be affected by the presence of D, C or S.

Comparison	Model
Attribute-wise	Multi-alternative decision field theory
	Leaky competing accumulator model
Alternative-wise	Comparison-grouping model
Attribute-and-alternative-wise	Decision by sampling
	2N-ary choice tree model
	Multi-alternative linear ballistic accumulator model

Table 5.1: List of models discussed

to Car A in both quality and economy dimensions and should thus be discarded but, after adding this decoy, Car A becomes more likely chosen and Car B becomes less likely chosen (Huber et al., 1982). Adding Car C to a choice between Cars A and B produces the compromise effect. Car C has extremely good quality but poor economy. Importantly, Car C makes Car A a compromise between the other cars, and with Car Cs presence, Car A becomes more likely to be chosen than Car B (Simonson, 1989). The similarity effect is produced by adding Car S instead. Car S is similar to Car B, and Car Ss introduction results in the higher probability of Car A being chosen than Car B (Tversky, 1972).

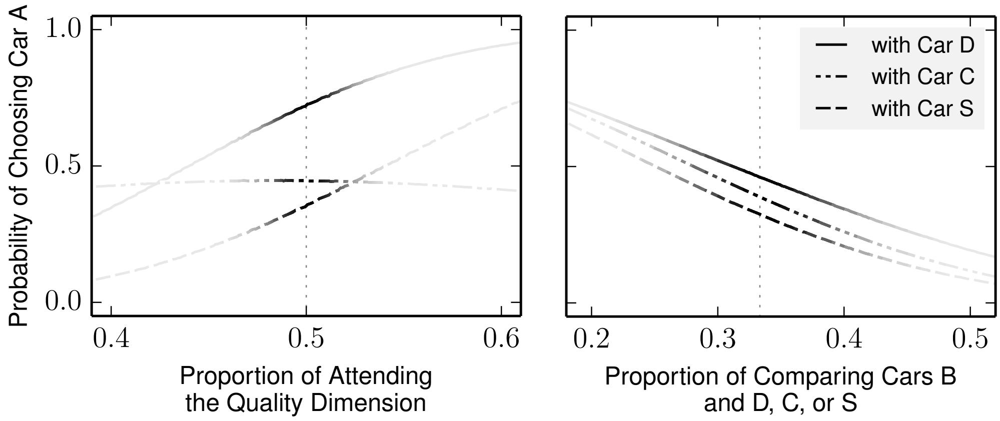
For the non-chosen alternative to influence a choice as described above, an individual has to be comparing alternatives in making a choice (e.g., Simonson, Bettman, Kramer, & Payne, 2013). Here we explore the nature of these comparisons, and consider models involving attribute-wise comparison, alternative-wise comparison, and attribute-and-alternative-wise comparison (see Table 5.1 for a list of the models).

According to attribute-wise comparison models, one attribute dimension is attended at one moment and all the available alternatives are simultaneously evaluated. In the above car example, an individual may attend, for instance, on the quality dimension of available cars at one moment, and evaluate how advantageous each of the three cars is. Then at the next moment, the individual may attend the economy dimension and evaluate all three cars. This attribute-wise comparison is implemented in multi-alternative decision field theory (Roe et al., 2001) and the leaky competing accumulator model (Usher & McClelland, 2001) to explain the three context effects.

In contrast, alternative-wise comparison models assume that all the attributes are integrated before comparison: one pair of alternatives is attended, attribute dimensions are integrated within each alternative, and then the pair of alternatives are compared on their integrated values. In the above example, an individual may integrate the quality and economy dimensions for, for instance, Car A, and also integrate these dimensions, separately, for Car B. Then, the individual compares the integrated value for Car A with Car B. At the next moment, the individual may select a new pair of alternatives, Cars A and S, and repeat the integrate-then-compare process. This integration of information across attributes is commonly assumed in models of two-alternative choice, including models where risk and reward information are integrated into a single expected-value-like measure such as cumulative prospect theory (Tversky & Kahneman, 1992) and the transfer of attention exchange model (Birnbaum, 2008). In the domain of multi-alternative choice, the comparison-grouping model (Tsuzuki & Guo, 2004) implements a mixture of attribute-wise and alternative-wise comparisons to explain the context effects.

Lastly in the attribute-and-alternative-wise comparison, one attribute dimension and also one pair of alternatives are attended at one moment, and two alternatives are compared against each other on the attended attribute dimension. For instance, an individual may attend on the quality dimension and compare Cars A and B at one moment. Then, at the next moment, the individual may focus on the economy dimension and compare Cars A and D. This comparison is assumed in the decision by sampling model (Stewart et al., 2006), which has been applied to context effects in risky and intertemporal choice (Stewart, Reimers, & Harris, in press) and could potentially be extended to account for the three context effects. The attribute-and-alternative-wise comparison has also been employed in the 2N-ary choice tree model (Wollschläger & Diederich, 2012), and the multi-attribute linear ballistic accumulator model (Trueblood, Brown, & Heathcote, in press).

This study examined predictions made by the three types of comparison model. In particular, we tested predictions concerning transitions of attention during choice and effect of random fluctuations in the attention on choice.



(a) Attribute-wise Comparison:
A Decision Field Theory Simulation

(b) Alternative-wise Comparison:
A Comparison Grouping Model Simulation

Figure 5.2: Simulated choice probability. The darkness of the line corresponds to the likelihood of the attention frequency given the equal weighting, and vertical dotted line represents most likely attention split. The left panel illustrates that when Car D or S is included in a choice set, more sampling of the quality dimension predicts higher probability of Car A to be chosen. The right panel shows that the probability of choosing Car A decreases with the frequency of comparisons between Cars B and D, C, or S.

5.1.1 The pattern of attention transition

In attribute-wise comparison, all of the available alternatives are simultaneously compared on a single attribute dimension. Therefore, an individual is likely to fix attention to one attribute dimension and shift their attention back and forth between alternatives to make comparisons. Thus we should see transitions of attention between alternatives within a single attribute dimension more frequently than, or at least equally frequently to, transitions within a single alternative between attribute dimensions. This same pattern of transitions is predicted by the attribute-and-alternative-wise comparison.

In contrast in the alternative-wise comparison models, all the attributes are used simultaneously in each comparison. Therefore, an individual is likely to fix attention to one alternative, shift their attention within the alternative to integrate attribute values, and then make a comparison. Thus we should see transitions of attention between attributes within a single alternative more frequently than, or at least equally frequently to, between alternatives.

5.1.2 The influence of stochastic fluctuations in attention on choice

When attribute dimensions are weighted equally so that each attribute dimension is equally likely to be attended at any moment, there will still be trial-to-trial variations in the number of times each attribute dimension is attended. This is due to the stochastic nature of the allocation of attention, and the relative frequencies of the observed split in attention are given by the binomial distribution. For example, with two equally weighted dimensions and with 10 allocations, the number of times each dimension is attended would follow the binomial distribution. So we would see a 5/5 split 24.6% of the time but, just by random chance, we would see the unequal splits (0/10, 1/9, 2/8, 3/7, or 4/6) 75.4% of the time.

Thus, for a particular trial, one attribute dimension will often be attended more frequently than another, even when attribute dimensions are weighted equally. These trial-by-trial fluctuations will increase the probability of selecting the alternatives high on the more attended dimension. To illustrate this prediction, we simulated the multi-alternative decision field theory. In this simulation, a choice is reached after 1,000 comparisons and dimensions are weighted equally. We explored how the choice probabilities change with the number of times the quality dimension is attended. The results are summarized in the left panel in Figure 5.2 (see Appendix A.1 for the details). This figure illustrates that, for example, when 490 comparisons are made on the economy dimension and 510 comparisons are on the quality dimension, probability of choosing Car A is .69 with the presence of Car D in the choice set. Generally when Car D or S is included in a choice set, more sampling of the quality dimension predicts higher probability of Car A to be chosen.

We also considered attention fluctuating over pairs of alternatives in the alternative-wise comparison models. One pair of alternatives will more frequently be compared against each other even with an equal weighting of all the pairs. This stochastic bias towards one pair of alternatives results in these alternatives being more likely to be chosen. For example if an individual more frequently compares Cars B and C, the individual is more likely to choose Car B or C and less likely to choose Car A. To illustrate this prediction, we simulated a modified version of the comparison grouping model. This modified version assumes that all the pairs of alternatives are equally weighted and that an alternative is chosen after 1,000 alternative-wise comparisons (see Appendix A.2 for the details). We manipulated the number of comparisons made between Cars B and D, C, or S, and summarized the results in the right panel of Figure 5.2. The figure shows that the probability of choosing Car A decreases with the frequency of comparisons between Cars B and D, C, or S.

Finally, in the attribute-and-alternative-wise comparison models, bias towards one pair of alternatives affects choice. But here, attention to an alternative pair interacts with attention towards an attribute dimension: In the car example, an individual is more likely to choose Car A over B if the individual more frequently compares Cars A and B on the economy dimension. In contrast, frequent comparison of Cars A and B on the quality dimension should lead to the choice of Car B.

In summary, the three types of model predict different patterns of attention transition and make competing claims on how a stochastic attention bias explains choice. These claims were examined using eye-movement data in the following experiment.

5.2 Method

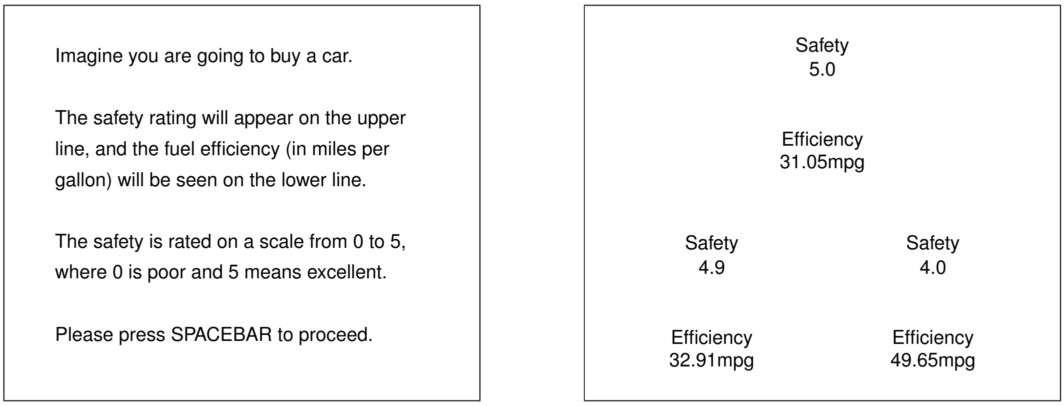
Following Simmons, Nelson, and Simonsohn (2012)s recommendation, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

5.2.1 Participants

One hundred undergraduate students were recruited through the participant panel at the University of Warwick and were paid £5.00 for participating. We decided in advance of collecting the data to test exactly 100 participants. Our previous work indicated that this would give us reasonable statistical power to replicate the attraction, compromise, and similarity effects. Seven participants could not complete the experiment due to failure in tracking their eyes (e.g., lazy eyes), leaving 93 (34 males and 59 females) participants. Their ages ranged from 17 to 49 (median = 21.0).

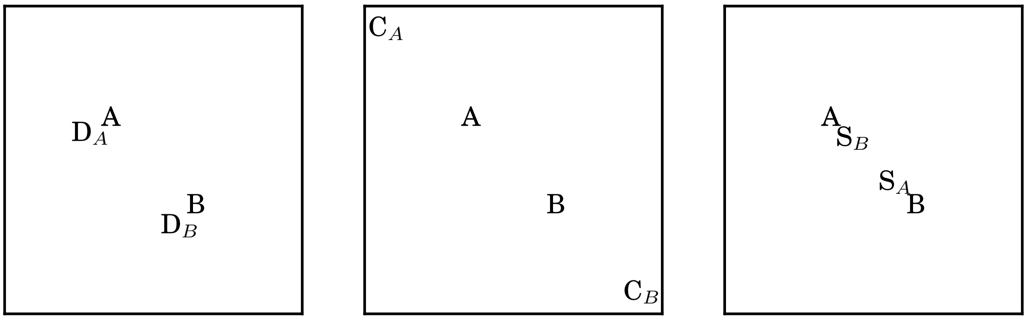
5.2.2 Procedure

Participants made 40 choices. At the beginning of each choice, participants were given information about the two attributes involved. After displaying the fixation point until participant fixated it for at least 500 ms, the experiment program presented three choice alternatives: one at the lower left corner of the screen, another at the top middle area, and the other at the lower right corner of the screen. This presentation ensures that the three alternatives are equally distant from each other on the display. An example screen shot is given in Figure 5.3. Participants made a



(a) Choice Description (b) Choice Set

Figure 5.3: Example screen-shots. This example depicts a choice between cars in an attraction choice. Font size is enlarged for this illustration.



(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.4: Locations of the alternatives used in the experiment.

choice by pressing one of the left, up, or right arrow keys.

The 40 choices comprised 10 attraction, 10 compromise, 10 similarity, and 10 catch choices. The catch choices always had one dominant alternative on both attribute dimensions. We used participants responses to the catch choices to assess whether they were engaged in the task. Each of the other 30 choices appeared in one of two versions, one favoring Alternative A and another favoring B. The two versions are summarized in Figure 5.4.

The left panel in Figure 5.4 displays the alternatives for the two versions of the attraction choices. One version involved Alternatives A, B and D_A (the decoy to Alternative A), and the other involved Alternatives A, B, and D_B (the decoy to Alternative B). The middle panel displays the alternatives for the compromise choices. One version involved Alternatives A, B, and C_A (making Alternative A the compromise), and the other involved Alternatives A, B, and C_B (making Alternative B the compromise). The right panel displays alternatives for the similarity choices: one version involved Alternatives A, B, and S_A (adding an alternative similar to Alternative B), and the other involved Alternatives A, B, and S_B (adding an alternative similar to Alternative A). The allocation of versions was counterbalanced between participants.

Each of 40 choices involved a different cover story (e.g., cars, laptops, and TV sets), and the same cover story was used for the two version of choices. Thus, all the participant made a choice between cars in an attraction choice, regardless of the version to which they were assigned. The order of the choices was randomized and the four types of choices were interleaved. The locations of alternatives and attributes on the screen were randomized for each choice.

Throughout the experiment, participants eye-movements were recorded at 500 Hz using an EyeLink 1000 (SR Research). The eye-tracker was placed right under the 19 inch monitor, and the distance between participants eye and the eye-tracker was kept between 50cm and 55cm. Also, the eye- tracker was calibrated just before the experiment and also after every 10 choices during the experiment.

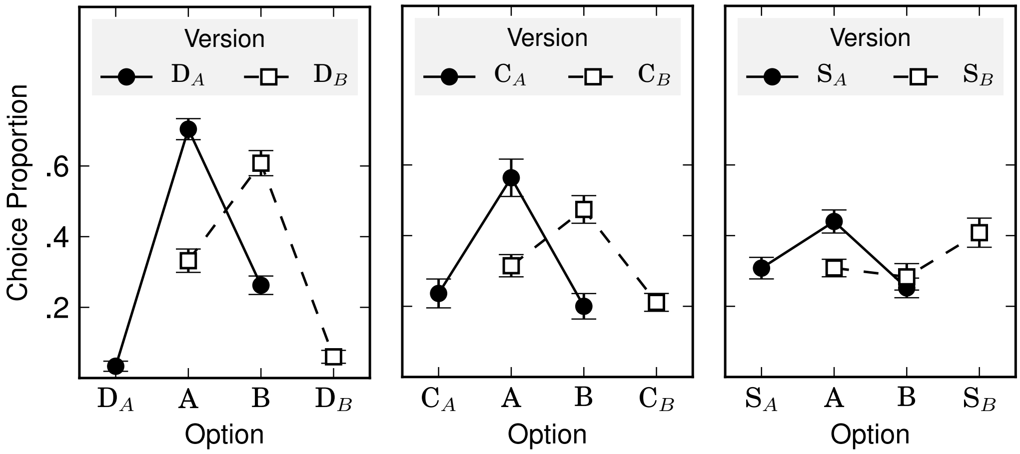
5.3 Results

Out of 93 participants, 44 participants did not choose the dominant alternative in one or more of the catch choices. These *less-engaged* participants may have been unable to differentiate the attraction and similarity choices, where the detection of dominance is crucial. Thus, the analysis below includes engagement as a factor, noting where it matters.

5.3.1 The attraction, compromise, and similarity effects were replicated

We computed the proportion of times each alternative was chosen for each choice type for each participant. The choice proportions from the engaged group of participants are plotted in Figure 5.5. The filled circles with the solid line represent the version of choices which favors Alternative A, and the empty squares with the dashed line represent the one which favors Alternative B.

The left panel in Figure 5.5 shows a replication of the attraction effect: Alternative A is most often selected from the D_A version while Alternative B is most frequently chosen from the D_B version. The middle panel shows a replication of the compromise effect: The compromise alternatives are most often to be chosen



(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.5: Mean choice proportions of the engaged participants. Error bars are standard error of the mean. In the attraction choices, D_A is inferior to A in both attributes, and D_B is inferior to B. In the compromise choices, C_A makes A a compromise between the other alternatives, and C_B makes B a compromise. In the similarity choices, S_A is similar to B and S_B is similar to A. The subscript of labels indicate which alternative (A or B) is favored by the context.

in both C_A and C_B versions. The right panel shows a replication of the similarity effect: Alternative A has a higher proportion of choice in the S_A version compared to the S_B version and Alternative B has a higher proportion, albeit only slightly, in the S_B version compared to the S_A version.

We explored the significance of these effects using a mixed-effect model. The proportions for Alternatives A and B are logit-transformed after multiplying with 0.9 to handle ones and adding 0.05 to handle zeros. Then the transformed proportions are entered into a linear mixed-effect linear model. The model had fixed effects for alternative (A or B), version (whether the version favors Alternative A or B), choice-type (attraction, compromise, or similarity), and participant group (engaged or less-engaged). The model also had by-participant slopes and intercepts as random factors.

The model fit indicates that the effect of the three-way interaction depends on participant engagement: the four-way interaction is significant, $\chi^2(2) = 17.92$, $p < .001$. When the same mode is fit only to the engaged group of participants, the three-way interaction effect indicates that the effect of choice alternative on the choice proportion depends on the choice type and the version: $\chi^2(2) = 28.84$, $p < .001$. Thus, we fit the same mixed-effect model to the attraction, compromise and similarity choices separately for the group of engaged participants.

For the attraction choices, the interaction effect is significant, $\chi^2(1) = 44.47$, $p < .001$, indicating that the choice proportions for Alternatives A and B are different between the D_A and D_B versions. The interaction effects are also significant for the versions. The interaction effects are also significant for the compromise choices, compromise choices, $\chi^2(1) = 19.90$, $p < .001$, and for the similarity choices, $\chi^2(1) = 4.11$, $p = .043$. These interaction effects indicate that the attraction, compromise and, similarity effects are replicated in this study.

For the group of the less-engaged participants, the three-way interaction is also significant: $\chi^2(2) = 51.81$, $p < .001$. The attraction and compromise effects are confirmed: $\chi^2(1) = 34.23$, $p < .001$ and $\chi^2(1) = 27.37$, $p < .001$. However, the similarity effect does not reach significance: $\chi^2(1) = 0.11$, $p = .738$.

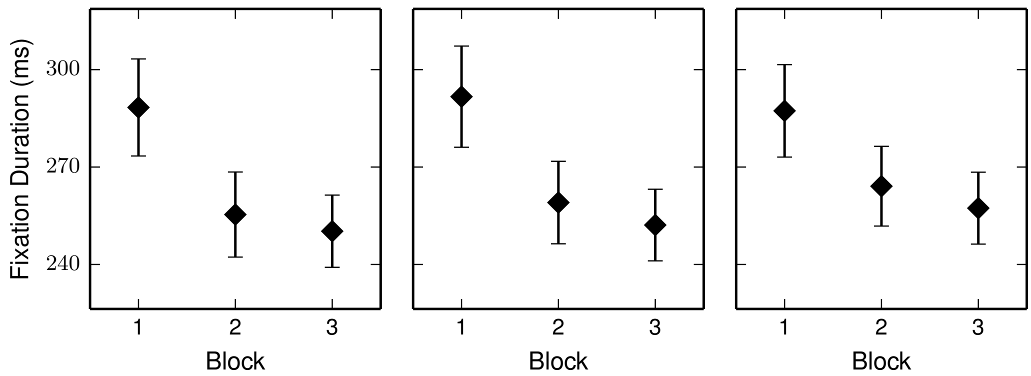
5.3.2 Eye Movements

For the fixation data, we defined non-overlapping regions of interest to identify to which alternative and attribute dimension the participant fixated his or her eye on. Due to noise in the detecting fixation locations, fixations were not registered for some of the displayed attributes in 153 out of the total of 2790 (= 30 choices $\times$ 93 participants) trials. These trials are removed from the analysis. Then, each choice was recoded to match the labels in Figure 5.1. So for the D_A version of the attraction choices, D_A was relabeled D. For the D_B version, D_B was relabeled D and labels for A and B were swapped. In addition, the attribute dimensions were switched when relabeling the alternatives in the D_B version. Similar relabeling was done for the compromise and similarity choices.

5.3.3 Stages of decision making

Previous studies which analyze eye-movement often assume three stages of decision making: initial screening, evaluation and comparison, and validation prior to making a choice (e.g., Glaholt & Reingold, 2011; Russo & Leclerc, 1994). Glöckner and Herbold (2011) review evidence that the duration of fixations increases with processing difficulty, and so differences in fixation duration over time may indicate different processing stages.

To examine the stages of decision making, we segmented the sequence of fixations into three blocks. Each block has the equal number of fixations, but when the number of fixations is not dividable by three, we added the reminder to the last block. Then the mean fixation duration is computed for each block for each



(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.6: Fixation duration as a function of time. Error bars are standard error of the mean.

participant and displayed in Figure 5.6. This figure illustrates that the fixation tends to be longer in the first block.

The fixation durations were examined with a mixed-effect model. Fixed effects are fixated alternative (A, B, or the third alternative), fixated attribute dimension, block (1, 2, or 3), and choice-type (attraction, compromise, or similarity). Random effects are by-participant intercept and slope for block. The interaction effects indicates that the fixation duration does not differ between the alternatives or the attribute dimensions ($ps > .066$), but that the difference between blocks depends on the choice-type ($\chi^2(4) = 13.48$, $p = .009$). Thus, we fit the mixed-effect model separately for each choice-type. Although the strength of the effect may differ between the choice-types, the effect of block is significant for all the three choice-types ($\chi^2(2) = 49.34$, $p < .001$ for the attraction choices; $\chi^2(2) = 44.92$, $p < .001$ for the compromise choices; $\chi^2(2) = 37.37$, $p < .001$ for the similarity choices), indicating that the fixation duration is significantly longer in the first block.

The longer fixation in the first block may indicate a qualitatively different stage of decision making. Therefore, we examined effects of the block in the following analysis, although the results hold if the block is not included in the analysis.

5.3.4 The pattern of attention transitions

According to the attribute-wise and attribute-and-alternative-wise comparison models, transitions of attention between alternatives on a single attribute should be more frequent than, or at least equally frequent to, transitions between attributes within a single alternative. In contrast, according to the alternative-wise comparison models,

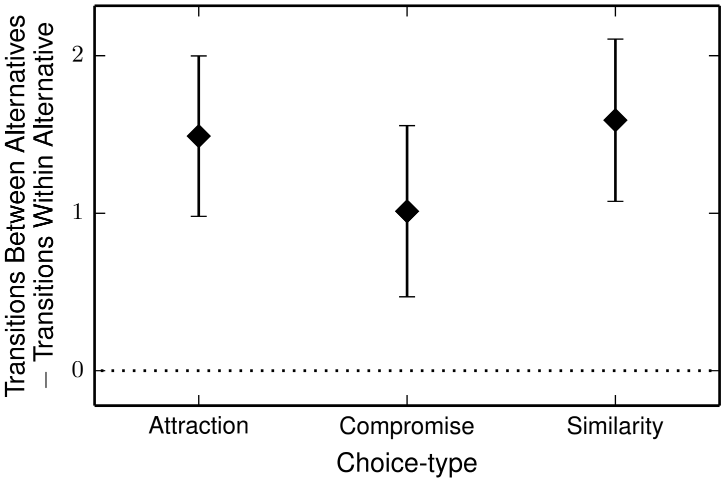


Figure 5.7: Difference in transitions. Error bar are standard error of the mean.

transitions of attention within a single alternative should be more frequent than, or at least equally frequent to, transitions between alternatives.

The difference between the number of transitions between alternatives and within an alternative is displayed in Figure 5.7. The between-alternatives transitions include only those on a single attribute dimension, excluding the between-alternatives, between-attributes transitions. Thus, the between-alternative transitions are underestimated, which should favor the prediction from the alternative-wise comparison. However, between-alternatives transitions are more frequently observed than within-alternative transitions, consistent with the attribute-wise and attribute-and-alternative-wise comparison models.

These transitions are examined with a mixed-effect model, whose fixed effects are participant group (engaged or less-engaged), choice-type (attraction, compromise, or similarity), and block (1, 2, or 3), and the random effects are by-participant slopes and intercepts. The three-way and two-way interaction effects indicate that the effect of transition-type does not differ significantly between the participant groups or the blocks: $ps > .064$. The main effect of choice-type is significant: $\chi^2(1) = 6.87$, $p = .032$. When the mixed-effect model was fit to each choice-type, the estimated intercept indicates the scores are significantly different from zero: $\beta = 0.48$ (95% CI [0.29, 0.68]) for the attraction choices; $\beta = 0.33$ (95% CI [0.12, 0.53]) for the compromise choices; $\beta = 0.52$ (95% CI [0.31, 0.73]) for the similarity choices. These results suggest that the effect differs quantitatively but not qualitatively across the choice-types.

Thus, the attention transitions more frequently between alternatives on the same attribute dimension than within an alternative. This pattern of attention transition supports the attribute-wise and attribute-and-alternative-wise comparison models and rejects the alternative-wise comparison models.

5.3.5 The influence of stochastic fluctuations in attention on choice

Before examining the influence of attention bias on choice — the subject of the simulations above — we need to consider the gaze-cascade effect (Shimojo, Simion, Shimojo, & Scheier, 2003). In the gaze cascade effect a developing preference for an alternative causes more frequent eye-movements to that alternative and more frequent eye-movements to an alternative causes preference for that alternative to develop, in a positive feedback loop. This gaze-cascade effect is considered independent of the comparison process and could artificially favor one prediction over another. Thus, we quantified the gaze-cascade effect and used it as a control variable.

Specifically, we counted the number of transitions *towards* an alternative and the number *away from* the alternative. According to Bird, Lauwereyns, and Crawford (2012), transitions towards an alternative increase the probability of selecting that alternative, and transitions away from an alternative decrease the probability of selecting that alternative. Importantly, once the number of transitions towards and away from an alternative is controlled for, there is no overall effect of the total number of fixations in predicting a choice.

As the number of transitions towards an alternative is highly correlated with the number of transitions away from the alternative, we took the difference as the gaze-cascade score. By definition, the number of transitions towards an alternative must be one less than, equal to, or one more than the number of transitions away from that alternative, and so the gaze-cascade score for an alternative was always -1, 0, or $+1$. When tested alone in a mixed-effect logistic regression with by-participant intercept and slope as random factors, the gaze-cascade score for Alternative A significantly predicts the choice of Alternative A: $\beta = 0.52$ (95% CI [0.45, 0.59]), $\chi^2(1) = 144.27$, $p < .001$. The gaze-cascade score for each alternative was entered as both fixed and by-participant random factors to all the models to predict choices we used below.

Attribute-wise comparison

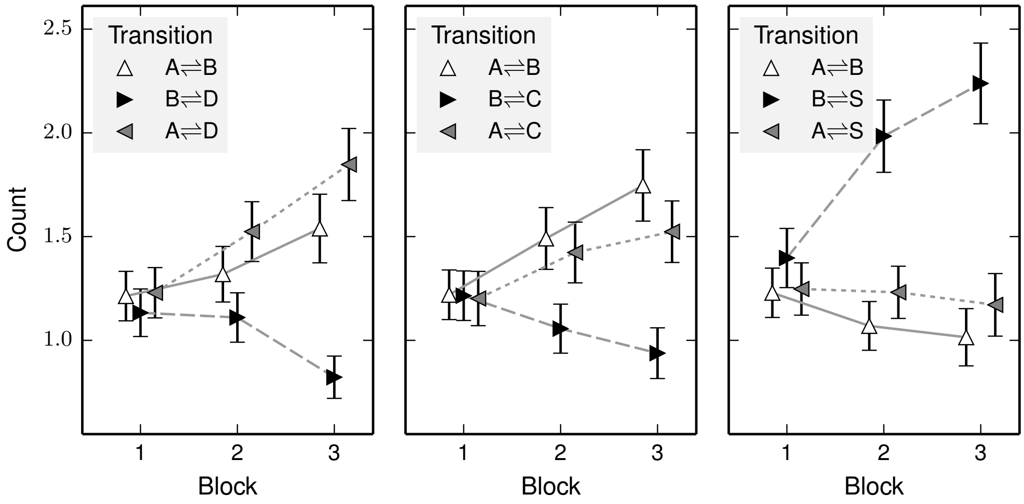
In the attribute-wise comparison models, a stochastic bias in attention towards one attribute dimension over the other should predict a choice of the alternative on which that attribute is highest, as in the simulation described above. To examine this prediction, we counted differences in the numbers of fixations and also summed the duration of fixations between the attribute dimensions within each trial. We first examined whether fixation counts and durations varied over the time course of a trial before testing whether fixation counts and durations were related to choices as the attribute-wise comparison models predict.

To examine whether these fixation counts and duration differ between the blocks, we used a mixed-effect model. Fixed factors are block (1, 2, or 3), participant group (engaged or less-engaged), and choice-type (attraction, compromise, or similarity), and random factors are by-participant slopes and intercepts. While we tested the counts and durations in the separate models, the total fixation duration is correlated with the fixation counts, as the average duration of each fixation does not differ significantly between alternatives or attributes (see the analysis in Section "Stages of decision making"). As a result, the model with fixation duration yielded essentially the same results as the model with the fixation counts. The model fits suggest that the counts and the duration do not differ significantly between blocks, participant groups or choice-types ($ps > .066$). Thus, we summed the counts and durations across the blocks to explore their relationship to choice.

The fixation counts and durations were then entered into mixed-effect logistic regressions to predict the choice of Alternative A. The fixed effects include participant group (engaged or less-engaged) and choice type (attraction, compromise, or similarity), and the random effects are by-participant slopes and intercept. The three-way and two-way interaction effects indicate that the effect of attention bias over attribute dimensions, in both counts and durations, does not depend on the participant group or the choice-type: $ps > .682$. The main effect suggests that attention bias is not a significant predictor of choice: $\beta = 0.00$ (95% CI [-0.01, 0.02]), $\chi^2(1) = 0.05$, $p = .821$ for the counts and $\beta = 0.00$ (95% CI [0.00, 0.00]), $\chi^2(1) = 2.02$, $p = .155$ for the duration. Thus, the prediction from the attribute-wise comparison models is not supported.

Alternative-wise comparison

In the alternative-wise comparison models, a bias in attention towards one pair of alternatives negatively correlates with probability of the remaining alternative being chosen, as in the simulation described above. To examine this prediction, we counted the number of transitions between each pair of alternatives within each trial. First we describe how, over the time course of a trial, some transitions come to be more frequent than others. Then we test whether the transition frequencies on a trial can



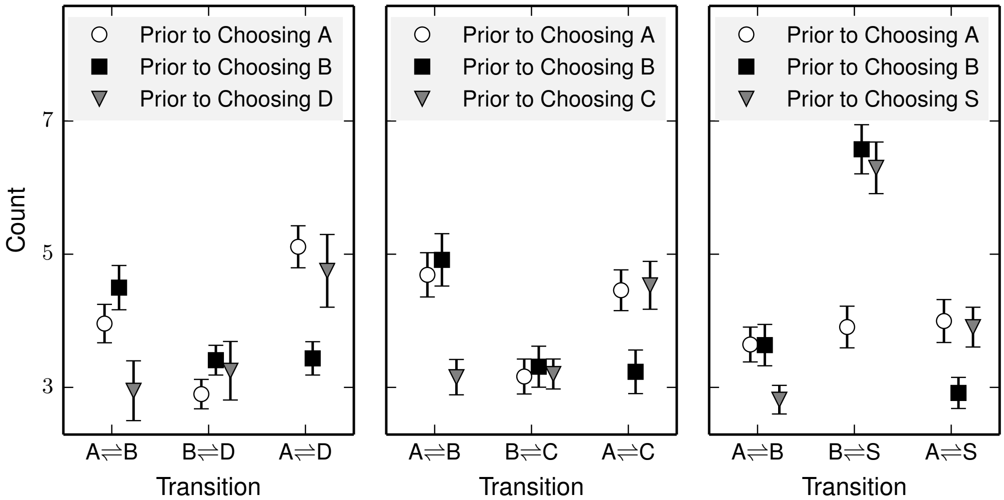
(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.8: Number of transitions prior to making a choice. Error bars are standard errors of the mean.

be used to predict the choice on that trial in the way alternative-wise comparison models predict.

The transitions are displayed in Figure 5.8 which shows that the differences between the transitions emerge over a trial. Transitions were transformed by adding 1 and logging before being entered into a mixed-effect linear regression in which the effects of block were examined. Fixed factors are block (1, 2, or 3), participant group (engaged or less-engaged), choice-type (attraction, compromise, or similarity), and transition (between A and B, between B and the third alternative, or between A and the third alternative), and the random factor was by-participant intercepts. Random factors do not include by-participant slopes, to keep the model complexity manageable. The four-way and three-way interaction effects indicate that the interaction effects do not depend on participant group ($ps > .308$), but that block has different effect depending on choice-type and transition ($\chi^2(8) = 385.89$, $p < .001$). Thus, the mixed-effect model is fit to each choice-type separately.

The model fits suggest that the effect of block depends on transition for all the choice-types (the attraction choices: $\chi^2(4) = 135.90$, $p < .001$; the compromise choices: $\chi^2(4) = 101.16$, $p < .001$; the similarity choices: $\chi^2(4) = 170.04$, $p < .001$). When the mixed-effect models were fit separately to each block, the model fits indicate that in the attraction and compromise choices, the transitions differ from each other non-significantly in Block 1 ($ps > .147$), but significantly in Blocks 2 and 3 (the attraction choices: $\chi^2(2) = 48.04$, $p < .001$ and $\chi^2(2) = 111.68$, $p < .001$,



(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.9: Number of transitions prior to making a choice. Error bars are standard errors of the mean.

respectively; the compromise choices: $\chi^2(2) = 35.60$, $p < .001$ and $\chi^2(2) = 87.38$, $p < .001$) In the similarity choices, the transitions differ in all the blocks (Block 1: $\chi^2(2) = 9.08$, $p = .011$; Block 2: $\chi^2(2) = 65.98$, $p < .001$; Block 3: $\chi^2(2) = 100.71$, $p < .001$).

Although the significance of the differences in the transitions differs between blocks, the direction of the differences is consistent across the blocks. In short, we see that, as the choice unfolds, differences in the frequencies of each transition type emerge. Before the attribute values have been read, there can be no bias to make some transitions more frequently than others. So gradually emerging differences in the transitions are entirely expected. The consistent directions of the differences imply that the process of decision making is not qualitatively different between the blocks.

The second state of our analysis is to see whether transition frequencies can be used to predict choices as the alternative-wise comparison models predict. We summed the transitions across the blocks and displayed the summed transitions in Figure 5.9. The transitions are largely consistent with the predictions for the alternative-wise comparison models for all three choice types. For example, the transition between Alternatives A and B is more frequent before A or B is chosen, and also the transition between Alternatives A and the third alternative is more frequent before A or the third alternative is chosen.

These transitions were entered into a mixed-effect logistic regression to predict the choice of Alternative A. The fixed effects include participant group and choice type, and the random effects are by- participant slopes and intercept. The interaction effects indicate that the effect of transitions does not significantly depend on the participant group ($ps > .567$), but that the effect depends on the choice-type: $\chi^2(6) = 82.53$, $p < .001$.

Consistent with the prediction, the transition between Alternatives B and D is a significant, negative predictor of choice A in the attraction choices: $\beta = -0.18$ (95% CI [-0.27, -0.09]), $\chi^2(1) = 16.11$, $p < .001$. Also the transition between B and S is a significant negative predictor in the similarity choices: $\beta = -0.39$ (95% CI [-0.48, -0.31]), $\chi^2(1) = 92.13$, $p < .001$. However, the effect of the transition between B and C is not significant in the compromise choices: $\beta = -0.02$ (95% CI [-0.10, 0.06]), $\chi^2(1) = 0.24$, $p = .626$.

In addition, some of the transitions, which involve Alternative A, significantly predict the choice of A. In the attraction choices, the transition between Alternatives A and D predicts the choice of A: $\beta = 0.27$ (95% CI [0.19, 0.35]), $\chi^2(1) = 49.37$, $p < .001$. This effect further implicates the alternative-wise comparison in the attraction choices, as the comparison between A and D always favors A.

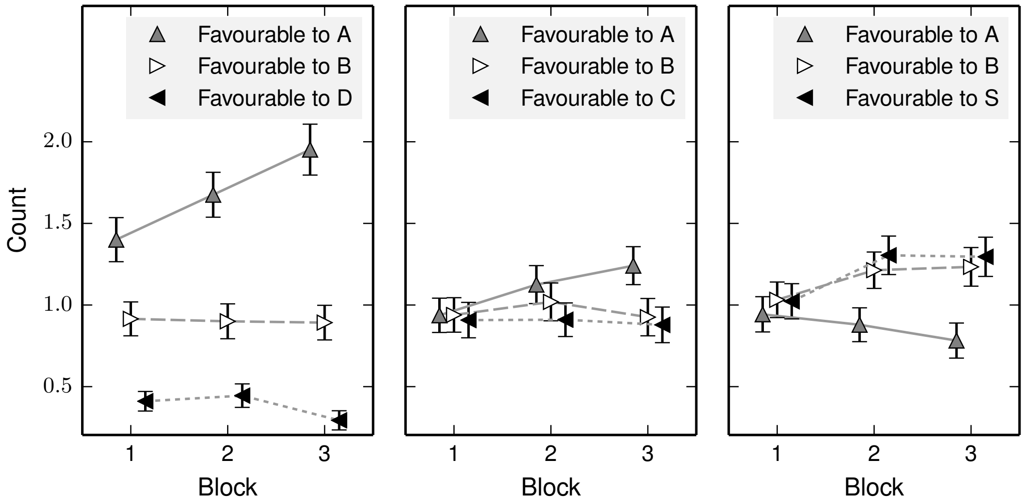
Also in the compromise choices, the transition between Alternatives A and B and also between A and C predict the choice of A: $\beta = 0.09$ (95% CI [0.02, 0.25]), $\chi^2(1) = 6.63$, $p = .010$; and $\beta = 0.13$ (95% CI [0.06, 0.20]), $\chi^2(1) = 12.11$, $p < .001$. In the similarity choices, the transition between Alternatives A and S predicts the choice of A: $\beta = 0.12$ (95% CI [0.04, 0.20]), $\chi^2(1) = 7.78$, $p = .005$.

The other significant predictor is not readily explained by the alternative-wise comparison. In the attraction choice, the transition between A and B negatively predict the choice of A: $\beta = -0.08$ (95% CI [-0.15, -0.01]), $\chi^2(1) = 4.69$, $p = .030$.

Thus, the effects of the transitions on choice are generally consistent with the predictions from the alternative-wise comparison, though some additional effects are not readily explained.

Attribute-and-alternative-wise comparison

According to the attribute-and-alternative-wise comparison, an attention bias towards one pair of alternatives predicts the choice of the alternative better on the attended attribute dimension, as described above. As the transitions are correlated between the attribute dimensions, we summed the numbers of fixation transitions favorable for each alternative, so that the larger count indicates more comparisons favorable to an alternative. For example, in an attraction choice, the transitions favorable for Alternative A is the sum of the transitions between Alternative A and B



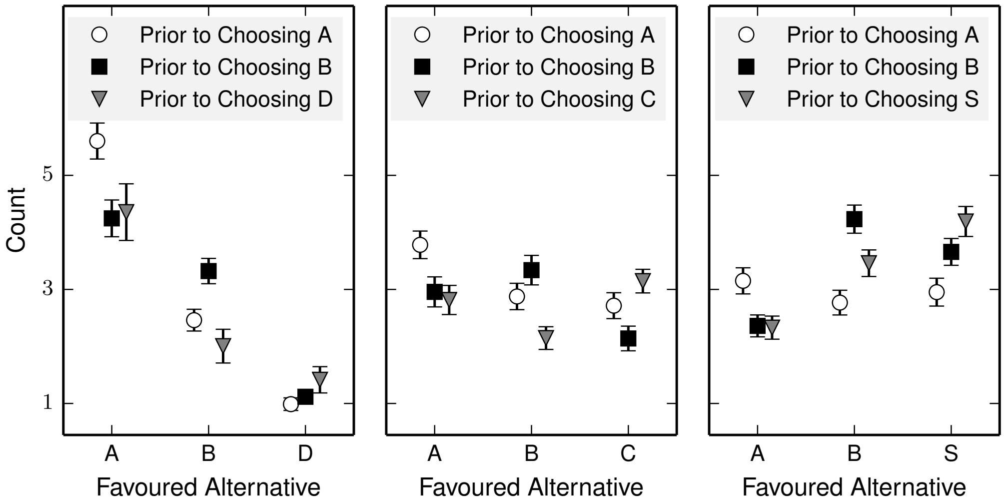
(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.10: Number of transitions prior to making a choice. Error bars are standard errors of the mean.

on the quality dimension and between A and D on both quality and economy dimensions. Our analysis here follows the same procedure as the analysis for attribute-wise comparisons and for alternative-wise comparisons above: First we explore how the counts of favorable transitions unfold over the time course of a trial and then we explore whether these counts predict choice as the attribute-and-alternative-wise comparison models predict.

The favorable transitions are summarized in Figure 5.10. In the attraction choices, the transitions favorable to Alternative A increase with block, because the transition between Alternatives A and D, which favors A, becomes more frequent with block (see Figure 5.8). Likewise in the compromise choices, as the transition between Alternatives A and B and between Alternatives A and C becomes more frequent, the transitions favorable to A increase. Also in the similarity choices, as the transition between Alternatives B and S becomes more frequent, the transitions favorable to B and S increase.

These transitions are transformed by adding 1 and logging, and then entered into a mixed-effect linear regression. Fixed effects are block (1, 2, or 3), participant group (engaged or less-engaged), choice-type (attraction, compromise, or similarity), and favored alternative (A, B, or the third alternative). By-subject intercepts are included in the random effects. The interaction and main effects indicate that the transitions do not differ significantly between participant groups ($ps > .430$), but that the effect of block differs between the choice-types and alternatives ($\chi^2(8) =$



(a) Attraction Choices (b) Compromise Choices (c) Similarity Choices

Figure 5.11: Number of transitions prior to making a choice. Error bars are standard errors of the mean.

167.29, $p < .001$). Thus, we fit the mixed-effect model to each choice-type separately.

In all three choice-types, the interaction and main effects indicate that the count does not differ between the participant groups ($ps > .097$), but that the effect of favored alternative depends on the block (the attraction choices: $\chi^2(4) = 107.95$, $p < .001$; the compromise choices: $\chi^2(4) = 37.30$, $p < .001$; the similarity choices: $\chi^2(4) = 55.36$, $p < .001$).

Again, although the significance of the difference differs between the blocks, the directions of the differences are consistent across the blocks. Thus, we summed the transitions across the blocks and displayed the transitions as a function of the chosen alternative in Figure 5.11.

Consistent with the prediction from the attribute-and-alternative wise comparison, the transition is generally higher for the alternative to be chosen. Before Alternative A is chosen (the empty circles), for example, the transition favoring A (the left column in each plot) is higher than the other transitions across the three choice-types.

These transitions were then entered into a mixed-effect logistic regression to predict the choice of Alternative A. The fixed effects include participant group and choice-type, and the random effects are by- participant slopes and intercept. The interaction effects indicate that the effect of transitions does not depend on the participant group ($ps > .28$), and that the effect depends on the choice-type: $\chi^2(6) = 39.25$, $p < .001$.

Consistent with the prediction, the transition, which indicates more frequent comparison favorable to Alternative A, is a significant predictor of choice A in the attraction, compromise, similarity choices: respectively, $\beta = 0.17$ (95% CI [0.11, 0.24]), $\chi^2(1) = 26.76$, $p < .001$; $\beta = 0.17$ (95% CI [0.07, 0.26]), $\chi^2(1) = 11.77$, $p < .001$; and $\beta = 0.24$ (95% CI [0.14, 0.34]), $\chi^2(1) = 20.69$, $p < .001$.

In addition, the transition, which indicates more frequent comparison favorable to other alternatives than A, is a significant, negative predictor of choice A in the attraction and similarity choices. In the attraction choices, the transition favorable to Alternative B shows $\beta = -0.19$ (95% CI [-0.29, -0.10]), $\chi^2(1) = 15.89$, $p < .001$. In the similarity choices, the transition favorable to Alternative B shows $\beta = -0.28$ (95% CI [-0.40, -0.17]), $\chi^2(1) = 25.18$, $p < .001$; and that favorable to S shows $\beta = -0.17$ (95% CI [-0.27, -0.08]), $\chi^2(1) = 12.51$, $p < .001$.

Thus, the results consistently support the prediction from the attribute-and-alternative-wise comparison.

5.4 Discussion

The present study replicated the attraction, compromise, and similarity effects in a within- participants design using 40 different consumer choice scenarios. This is the second study, following Berkowitsch et al. (in press), to simultaneously replicate the three effects in consumer choice. This is also the first study to record eye movements in participants showing these effects.

This study investigated the psychological processes of multi-alternative choice, focusing on how alternatives are compared against each other. We examined transitions of attention while a choice was being made. Specifically, the pattern of eye movements was examined to differentiate between three types of comparison model: attribute-wise, alternative-wise, and attribute-and-alternative-wise.

Transitions between alternatives within an attribute dimension were more frequent that transitions within alternatives between attributes, consistent with attribute-wise and attribute and alternative-wise models. The attribute-wise comparison models predict that bias towards a dimension should increase the probability that the alternative highest on that dimension should be chosen, but there was no significant effect of attribute-dimension bias. The alternative-wise comparison models predict that bias towards a pair of alternatives should decrease the probability that the third alternative is chosen, and this effect of alternative-pair bias was found. Finally, the attribute-and-alternative-wise comparison models alone predict an interaction between the alternative-pair and the attribute dimension attended. For a given pair and dimension, the alternative higher on the attribute dimension attended should be favored over the alternative lower on the attribute dimension attended. This interaction was observed. Overall, the eye movement data are most consistent with the attribute-and-alternative-wise comparison models, in which comparisons are between pairs of alternatives on single dimensions.

The finding of more transitions between alternatives within an attribute dimension could be influenced by physical locations of the alternatives within the display. In our experiment, the distance between the alternatives is deliberately very similar to the distance between two attribute values within an alternative. However, if between-alternative distances were minimized compared to within-alternative distances, attention might transition more frequently between alternatives, appearing as if the alternative- wise comparison is supported. Previous research however, has also favored the attribute-wise comparison over the alternative-wise comparison: When an individual is allowed to choose which information to examine, the individual more often decides to reveal information on one attribute dimension across available alternatives (Payne, 1976). Also, our results confirm previous findings, where transition is more frequent between alternatives on a single attribute dimension than within an alternative (Russo & Dosher, 1983).

While attribute-wise comparison models are supported by the attention-transition evidence, these models are not consistent with the null effect of attention bias on choice. The attribute-wise comparison models predict relationship between a choice and attention bias towards one attribute dimension over the other, but this prediction is not supported in our result. This result has implications for computational models beyond the class of comparison-based models described above. For instance, the associative accumulation model (Bhatia, 2013) explains the context effects with attention bias towards one attribute dimension over the other. Also, the range-normalization model (Soltani, De Martino, & Camerer, 2012) predicts that the attention bias should lead to different choices. These explanations are not consistent with the present results.

These relationships between transition and choices extend previous findings on eye-movement and choice. Krajbich and Rangel (2011) for instance, proposed a drift-diffusion model which incorporated fixations, by assuming that the drift rate was higher for fixated alternatives. This model relates a priori ratings of the attractiveness of alternatives and the fixation times on each alternative during a choice to the final choice of alternative. That is, to predict choice this model requires the pattern of fixations and also attractiveness judgments for each alternative. In contrast, our study focused on predicting choices from the pattern of attention transitions alone.

A second major finding is the gaze cascade effect (Shimojo et al., 2003; Simion & Shimojo, 2007). The gaze cascade effect is a developing bias to direct an eye-fixation toward the alternative ultimately chosen. As a result, choice can be predicted from transitions (Bird et al., 2012). While it is not clear how well this gaze cascade effect correlates with preference development (e.g., Glaholt & Reingold, 2009, 2011), our results confirm the positive relationship between the gaze cascade (measured from transitions) and choice. Also, the results show that even after controlling for this gaze cascade effect, the pattern of transition still predicts choices and reveals details of the comparison process in choice making.

In our analysis we were careful to explore how eye movements changed over the time course of a choice, because changes might indicate different stages of processing (e.g., Glöckner & Herbold, 2011). Although we did find a surprising tendency for early fixations to be longer, we did not find any qualitative shift in the pattern of eye movements. Instead, each of the biases we find emerged gradually within a trial. This is not surprising because, given that the biases are defined only by the relation of the attribute values to one another and not by some more obvious cue like physical location on the screen, these biases cannot emerge until the attribute values have been read and compared. Our data are most consistent with a single continuous cognitive process operating over the whole time course of a choice, unless attention is independent to eye-movements. Eye-movements are, however, tightly associated with attention (Kustov & Robinson, 1996), and its tight associations are also assumed in previous studies (e.g., Glöckner & Herbold, 2011; Krajbich & Rangel, 2011).

The effects of attention bias on a choice are consistent with the attribute-and-alternative-wise comparison models. Here a choice is reached through a series of comparisons of pairs of attributes on a single dimension, as in the decision by sampling model (Stewart, 2009; Stewart et al., in press, 2006; Stewart & Simpson, 2008), the 2N-ary choice tree model (Wollschläger & Diederich, 2012), and the multi-attribute linear ballistic accumulator model (Trueblood et al., in press).

Chapter 6

Multi-alternative Decision by Sampling

6.1 Background

As discussed in Chapter 5, the attraction, compromise, and similarity effects have been explained by various computational models. The majority of the models implement dynamic preference development over time, but differ from each other in two factors: how the model treats multiple attribute dimensions and how many alternatives are evaluated at one time. For instance, in decision field theory (Roe et al., 2001), the leaky competing accumulator model (Usher & McClelland, 2001) and the associative accumulation model (Bhatia, 2013), an individual attends one attribute dimensions at one time and simultaneously evaluates all the alternatives. In contrast, in the comparison grouping model (Tsuzuki & Guo, 2004) assumes that the attribute dimensions are integrated, but that an individual evaluates only two of the alternatives at a time.

Contrary to these processes implemented in the existing models, empirical evidence on choice process indicates that, at any moment, an individual attends to one attribute dimension and evaluates two alternatives (Payne, 1976; Russo & Leclerc, 1994; Russo & Dosher, 1983, Chapter 5). Based on the processes implicated by these empirical findings, we propose a new model of multi-alternative decision making: multi-alternative decision by sampling (MDbS). This new model implements three components of decision-making: 1) an alternative is evaluated through a series of pair-wise comparisons on single attribute values, 2) similar alternatives are compared more often, and 3) relatively small differences in attribute values are ignored.

We first review empirical findings to support each process and discuss how these processes explain the attraction, compromise, and similarity effects. Then the three processes are formulated in a computational model and, through simulations, we demonstrate that the MDdS model can produce the three context effects and other associated findings. The performance of the MDbS model is compared against decision field theory, and the last section addresses how the MDbS model relates to other existing models of decision making.

6.2 Qualitative Account of the Context Effects

In this section, we discuss empirical findings to support each of the three components and address how these components explain the attraction, compromise, and similarity effects. The attraction effect is explained by the first component, the compromise effect is explained by the first and second components, and the similarity effect is explained by the three components. First, we discuss the first component: pair-wise, single-attribute comparisons.

6.2.1 Pair-wise, single-attribute comparison explains the attraction effect

Pair-wise comparison has been reported in studies of eye-movement during choice. For example, when making a choice between more than two alternatives, an individual moves his or her eye back and forth between two of the alternatives most frequently (Russo & Leclerc, 1994). This finding indicates that a pair of alternatives is compared at one time. Also, Chapter 5 demonstrates that alternatives are compared on a single attribute dimension at a time.

To formulate these pair-wise, single-attribute comparison in a new model, we employ the framework of the decision by sampling (Stewart et al., 2006). Under this framework, an individual reaches a choice through a series of comparisons of pairs of attribute values on single dimensions (see Stewart & Simpson, 2008). These comparisons are between sample of attribute values in working memory. While some of the sample will come from long-term memories of attribute values, some of the working memory sample will come from the context given by the current choice set.

In our application to multi-alternative choice, we assume that an individual is unfamiliar with the attribute dimensions, and that this individual can only have samples from the current choice set. If an individual is familiar with the attribute dimensions, however, some of the working memory samples can come from long-term memory, and, as we discuss in the model simulation, the samples from long-

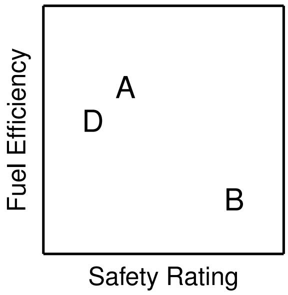


Figure 6.1: Attraction choice set.

Comparison	Attribute	Favored Car
A $\rightleftharpoons$ B	Safety Rating	B
A $\rightleftharpoons$ B	Fuel Efficiency	A
B $\rightleftharpoons$ D	Safety Rating	B
B $\rightleftharpoons$ D	Fuel Efficiency	D
D $\rightleftharpoons$ A	Safety Rating	A
D $\rightleftharpoons$ A	Fuel Efficiency	A

Table 6.1: List of comparisons for the attraction choice set.

term memory will dilute the attraction, compromise, and similarity effects. Exactly this dilution has been reported for the attraction (Kim & Hasher, 2005) and the compromise effects (Sheng, Parker, & Nakamoto, 2005).

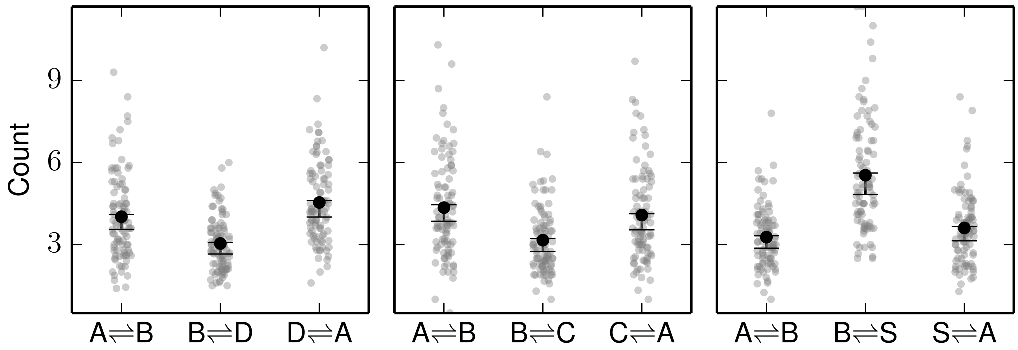
Thus, in the MDbS model, when deciding on a car to purchase, for example, an individual may compare Car A against Car B on fuel efficiency at one moment. In the next moment, the individual may compare Car A against Car D on the safety rating. The individual keeps track of the number of favorable comparisons for each car and ultimately chooses one car when sufficient comparisons are accumulated. This accumulation process stochastically approximate rank-position of alternatives (see Stewart et al., 2006).

This pair-wise, single-attribute comparison process explains the attraction effect. The choice set for the attraction effect is displayed in Figure 6.1, and the six possible comparisons of alternatives are listed in Table 6.1. Out of the six possible comparisons, three comparison favor Car A, while two comparisons favor Car B and one comparison favor Car D. Thus, Car A is most frequently favored, and this results in Car A having the largest choice probability.

6.2.2 Similarity bias explains the compromise effect

When comparing alternatives, an individual shifts their attention back and forth between two alternatives which share an attribute value (Russo & Rosen, 1975). Extending this finding, and to examine whether attention more frequently goes back and forth between a pair of similar alternatives than a pair of dissimilar alternatives, we re-analyzed the eye-movement data from Chapter 5.

Figure 6.2 displays the number of transitions between alternatives made before making a choice. These counts are examined with a Bayesian mixed-effect Poisson regression. The fixed factors are choice set (attraction, compromise, or sim-



(a) Attraction Choice Set. (b) Compromise Choice Set. (c) Similarity Choice Set.

Figure 6.2: Counts of eye-fixation transition between alternatives. Grey circles, jittered randomly along the horizontal axis, represent by-participant means, and black circles indicate grand means. The error bars show the 95% highest density interval of the mean.

ilarity) and transition (between A and B, between B and T, and between A and T). We label the other alternative T, where T equals D in the attraction choice sets, C in the compromise choice sets, and S in the similarity choice sets. The random factors are by-participant slopes and intercepts. For all the fixed and random factors, we used a non-informative prior[1] and drew 10^5 samples from the posterior using the no-u-turn sampler (Hoffman & Gelman, 2014). We discarded the first half and pooled the remaining samples from four chains without thinning.

Estimated coefficients in Poisson regressions indicate estimated difference in ratio, and the interaction coefficients indicate that the ratio of A⇌B to B⇌T counts differs across the attraction and similarity choice sets: posterior probability distribution supports the difference with a probability $p_{posterior} > .999$. The A⇌B to A⇌T ratio also differs between the attraction and compromise choice sets: $p_{posterior} > .999$. Thus, we kept the interaction term in the model and examined the posterior estimates. The 95% highest density intervals in Figure 6.2 show that given the data, we are 95% confident that the population mean falls within the interval. Below, we report where the intervals do not overlap and the difference is greater than 1.10 in ratio at the posterior median.

In the attraction choice set, the mean transition count between A and D is 1.13 times greater than between A and B and also 1.50 times greater than that between B and D. In addition, the mean count between A and B is 1.32 times greater than that between B and D. In the compromise choice set, the mean transition count

[1]The prior for a fixed factor is $\mathcal{N}(0, 1000)$ and the prior for a random factor is $\mathcal{U}(0, 1000)$.

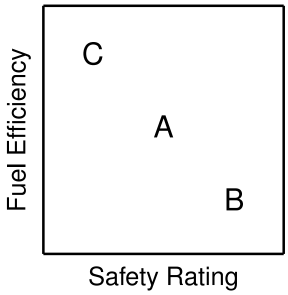


Figure 6.3: Compromise choice set.

Comparison	Attribute	Favored Car
A ⇌ B	**Safety Rating**	**B**
A ⇌ B	**Fuel Efficiency**	**A**
B ⇌ C	Safety Rating	B
B ⇌ C	Fuel Efficiency	C
C ⇌ A	**Safety Rating**	**A**
C ⇌ A	**Fuel Efficiency**	**C**

Table 6.2: List of comparisons for the compromise choice set.

between A and B is 1.40 times greater than that between B and C. Also, the mean count between C and A are 1.29 times greater than that between B and C. Lastly in the similarity choice set, the mean count between B and S is 1.67 times greater than that between A and B and also 1.53 times greater than that between S and A.

These results can be summarized by stating that the attention goes back and forth more frequently between similar alternatives. This similarity-biased comparison, together with the pair-wise, single-attribute comparison, explains the compromise effect. The six possible comparisons in the compromise choice set are listed in Table 6.2. The rows in bold font in this table represent the comparisons of similar alternatives. Both comparisons in which Car A is favored are frequent, but only one of the two for Car B and one of the two fro Car C is frequent. Thus Car A will be favored.

6.2.3 Just meaningful difference explains the similarity effect

When making a comparison, an individual may not register a relatively small difference. We borrow the notion of the *just noticeable difference* from studies on sensory discrimination. When the difference between two sensory stimuli (e.g., loudness and brightness) is too small, the stimuli are often not recognized as different. To see a stimulus as different from another, the ratio of their magnitudes is required to be above a certain value, called the Weber function.

When applied to decision making, the just noticeable difference is often called *just meaningful difference.* In support for the just meaningful difference, Tversky and Kahneman (1981) report that an individual is willing to make an extra trip to save $5 on a $15 purchase but unwilling to make the same trip to save $5 on a $125 purchase. This finding suggests that the discount is judged not on an absolute difference but on an ratio. Although the reduced monetary amount is $5 in both

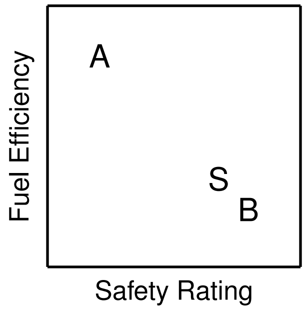


Figure 6.4: Similarity choice set.

Comparison	Attribute	Favored Car
A $\rightleftharpoons$ B	Safety Rating	B
A $\rightleftharpoons$ B	Fuel Efficiency	A
B $\rightleftharpoons$ S	**Safety Rating**	**B**
B $\rightleftharpoons$ S	**Fuel Efficiency**	**S**
S $\rightleftharpoons$ A	Safety Rating	S
S $\rightleftharpoons$ A	Fuel Efficiency	A

Table 6.3: List of comparisons for the similarity choice set.

cases, the $5 discount is 33% reduction from the price of $15 but is only 4% reduction from the price of $125. The 4% reduction may not be a meaningful enough difference to influence a decision.

The just meaningful difference is also reported in the studies of price perception, which report that a relatively small change in prices, expressed in percentages, does not have a significant impact on consumer choices (Kalwani & Yim, 1992). Also, studies on employees' judgments of increases in salary report that the increment expressed in percentage is a better predictor of employees' judgments on meaningfulness of the increment (Heneman III & Ellis, 1982; Futrell & Varadarajan, 1985) and also employees' subsequent spending and saving decisions (Rambo & Pinto, 1989).

The just meaningful difference, together with the pair-wise, single-attribute comparison and similarity bias, can explain the similarity effect. The possible comparisons within the similarity choice set are listed in Table 6.3. The shaded rows indicate comparisons discounted because the differences are small. Now, Cars B and S are so similar to each other that the difference is unlikely to be meaningful. As a result, the comparison between B and S is discounted. Outside this discounted comparison, Car A is most frequently favored, resulting in the highest probability that Car A is chosen among the three cars.

6.2.4 Differences from the decision by sampling model

Thus far, we have argued that the MDbS model accounts for the attraction, compromise and similarity effects, with only the additional component: comparisons are more likely between similar pairs (component 2 above). The other two components were both core components in the original decision by sampling model (e.g., see Stewart & Simpson, 2008).

6.3 Quantitative Specification

Now that we have qualitatively described the model, the mathematical formulation will be presented so it can be applied to the data. The MDbS model keeps track of the frequency of comparisons favoring each alternative, and predicts the choice of the first alternative whose frequency count reaches the decision threshold, λ. The three cars in a choice set are labeled as Cars A, B, and T, where Car T represents Car D, C, or S. Comparison of Car A against Car B in working memory is denoted as $A \rightharpoonup B$, and the probability of this comparison is $p(A \rightharpoonup B)$. Also, $A_i \not\equiv B_i$ denotes registration of difference between Car A and Car B on attribute dimension i. Thus, $p(A_i \not\equiv B_i)$ represents the probability of recognizing the difference as meaningful.

6.3.1 Pair-wise, single-attribute comparison

The probability that the frequency count for Car A increases at a given moment is defined as follows:

$$\begin{aligned} p(\Psi_A\text{++}) = p(A \rightharpoonup B) &\sum_i p(\text{Attend } i)\, p(A_i \not\equiv B_i)\, 1_{A_i \geq B_i} \\ + \; p(A \rightharpoonup T) &\sum_i p(\text{Attend } i)\, p(A_i \not\equiv T_i)\, 1_{A_i \geq T_i}. \end{aligned} \tag{6.1}$$

Here, i indexes attribute dimensions: A_i is the value of Car A at Attribute i. Also, $1_{condition}$ is an indicator function, whose value is 1 if the condition is met otherwise 0.

Thus, the probability of incrementing the number of favorable comparisons for Car A is a product of three factors: the probability of comparison, the probability of attending to one dimension, and the probability of registering the difference as meaningful. Previous research indicates that the probability of attending to an attribute dimension is not likely to be different across the choice sets (e.g., Bonaccio & Reeve, 2006). Thus we assume that $p(\text{Attend } i)$ is uniform across dimensions: 1 divided by the number of dimensions.

6.3.2 Comparison probability

The similarity between alternatives is computed with an interdimensional similarity function (Nosofsky, 1986; Shepard, 1987). Thus the similarity between Cars A and

B is computed as follows:

$$\eta_{A,B} = \exp\left(-\alpha\,\left(\sum_i \Delta_{A_i,B_i}\right)^{1/\gamma}\right)$$

Here, parameters α and γ control how the similarity is perceived: α controls overall discriminability in the attribute space, where the larger value of α indicates the better discriminability, and γ determines the shape of the similarity function (for a discussion, see Jäkel, Schölkopf, & Wichmann, 2008).

This symmetrical similarity is a function of the psychological difference between A and B on dimension i, denoted as $\Delta_{A_i,\,B_i}$. Following Weber's law, the difference is computed as a ratio:

$$\Delta_{A_i,B_i} = \left(\frac{\max(A_i,\ B_i)}{\min(A_i,\ B_i)}\right)^{\gamma} \tag{6.2}$$

Then, the probability of comparing Car A against Car B is defined as:

$$p(A \rightharpoonup B) = \frac{\eta_{A,B}}{\eta_{A,B} + \eta_{B,A} + \eta_{B,T} + \eta_{T,B} + \eta_{A,T} + \eta_{T,A}},$$

where the denominator normalises by dividing by the sum of all directional pair-wise similarities.

6.3.3 Just meaningful difference

Using the psychological difference defined in Equation 6.2, the probability of recognizing the difference between A_i and B_i as meaningful is computed with the logistic function:

$$p(A_i \not\equiv B_i) = \frac{1}{1 + \exp(\beta_0 + \beta_1\,\Delta_{A_i,B_i})},$$

where parameters β_0 and β_1 determines how large the difference should be to be recognised as meaningful.

6.3.4 Choice probability

The choice probability for Car A is equal to the probability that Ψ_A reaches threshold λ before Ψ_B or Ψ_T. As some comparisons do not result in an increment to either Ψ_A, Ψ_B, or Ψ_T, we normalize $p(\Psi_A\texttt{++})$ to derive $p^*(\Psi_A\texttt{++})$, the probability of an increment to Ψ_A, given there is an increment:

$$p^*(\Psi_A\texttt{++}) = \frac{p(\Psi_A\texttt{++})}{p(\Psi_A\texttt{++}) + p(\Psi_B\texttt{++}) + p(\Psi_T\texttt{++})}.$$

Then, the choice probability is computed as follows:

$$p(\text{Choose A}) = p^*(\Psi_A\texttt{++}) \sum_{\psi_B=0}^{\lambda-1} \sum_{\psi_T=0}^{\lambda-1} \mathcal{M}([\lambda-1, \psi_B, \psi_T] \mid p^*, \lambda - 1 + \psi_B + \psi_T),$$

where

$$p^* = [p^*(\Psi_A\texttt{++}), p^*(\Psi_B\texttt{++}), p^*(\Psi_T\texttt{++})].$$

Here, ψ_B indicates the value of Ψ_B when Ψ_A reaches the threshold, and $\mathcal{M}$ is the multinomial probability mass function: $\mathcal{M}([\psi_A, \psi_B, \psi_T] \mid [p^*(\Psi_A\texttt{++}), p^*(\Psi_B\texttt{++}), p^*(\Psi_T\texttt{++})], N)$ is the probability of Car A accumulating ψ_A favorable comparisons, Car B accumulating ψ_B comparisons and Car T accumulating ψ_T favorable comparisons, given the normalized probabilities, $p^*(\Psi_A\texttt{++})$, $p^*(\Psi_B\texttt{++})$ and $p^*(\Psi_T\texttt{++})$, and the total number of comparisons, N. Previously, the decision by sampling model used the binomial probability mass function (Stewart et al., 2006) to explain choice between two alternatives, but to explain multi-alternative choice, we use the multinomial function.

As the choice is made when Ψ_A reaches λ, the above computation ensures that the last comparison favors Car A. The probability of choosing Car B or T is computed in a similar manner.

6.4 Model Prediction

This section demonstrates the context effects with the above specified computational model. We first estimate parameter values using the multi-alternative choice data from Chapter 5, which replicated the attraction, compromise, and similarity effects. Specifically, we pooled the choice response data from all the participants who correctly selected the dominant alternative in each catch trial, and sampled parameters from the posterior distribution using the adaptive Metropolis sampler (Rosenthal, 2011). The prior distribution for all the parameters is a non-informative, improper distribution: the normal distribution with mean 0 and infinite variance.

We drew 10^6 posterior parameter values, discarded the first half, and thinned at an equal interval to retain 10^4 parameter values. At the median, parameter values are: $\alpha = 0.056$, $\beta_0 = 15.01$, $\beta_1 = -14.65$ $\gamma = 0.36$, and $\lambda = 3$.

Using these parameter values, the context effects were simulated on a hypothetical choice set. The choice set to simulate the attraction effect is illustrated in Figure 6.5a. These attribute values ensure indifference when only Cars A and B are available. Car A is 1.5 times better than B in the safety rating, and B is 1.5 times better than A in the fuel efficiency. Thus, probability of recognizing the difference as meaningful is the same for both dimensions. Also, please note that, given Equation 6.2, our model is invariant to unit changes. The following prediction results are identical if, for example, we use kilometers per liter for fuel efficiency.

In the following, we estimated the three context effects and also various empirical findings associated with these effects. As described above, the parameter values are constrained only on the choice responses on the choice sets from Chapter 5 which did not include the hypothetical choice set. Thus, the parameter values are not constrained to produce particular comparison probability or probability to recognize a difference as meaningful.

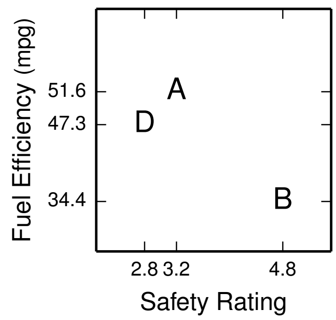
6.4.1 The three context effects

Attraction effect

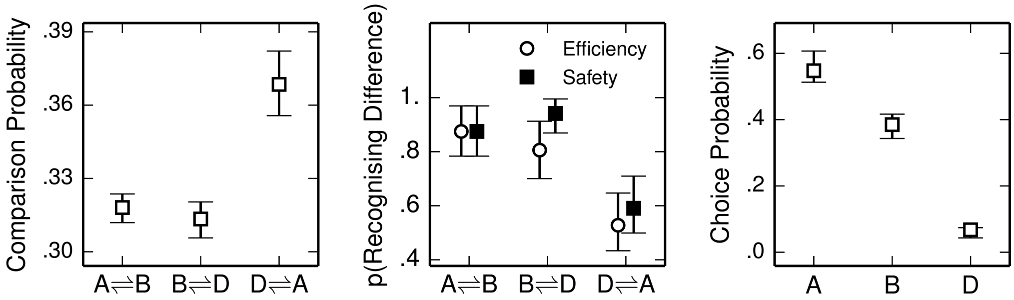
The comparison and choice probability were predicted on the choice set illustrated in Figure 6.5a. We summed $p(A \rightharpoonup B)$ and $p(B \rightharpoonup A)$ to get the comparison probability $p(A \rightleftharpoons B)$ in Figure 6.5b. This figure shows that, with Car D present, Cars A and D are most frequently compared against each other. The error bars here show that given the choice response data from Chapter 5, we are 95% confident that the comparison probability in this particular choice set falls into the interval. As the frequent comparison between Cars A and D favors Car A, Car A becomes most likely chosen, as displayed in Figure 6.5d.

The attraction effect is reported to be weaker when the decoy is closer to Car A (Soltani et al., 2012). To simulate this finding, we moved Car D from the point marked as E in Figure 6.6a toward where Car A is, and computed the probability of choosing Cars A and B. As Car D comes closer to Car B, the two alternatives become similar to each other. This similarity increases the probability of A$\rightleftharpoons$B comparison (Figure 6.6b), but reduces the probability that the difference is registered as meaningful (Figure 6.6c). When the difference between Cars A and D is not registered, preference for Car A becomes less likely to increase, resulting in the smaller difference in the probability that Car A is chosen compared to the probability that Car B is chosen (Figure 6.6d), just as Soltani et al. (2012) found empirically.

Previous research also reports that the strength of the attraction effect can



(a) Attraction choice set



(b) Comparison probability (c) Just meaningful difference (d) Choice probability

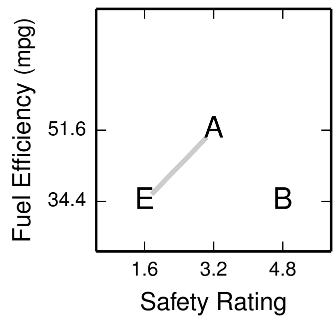
Figure 6.5: MDbS prediction of the attraction effect. Square and circle markers show prediction with the posterior median parameters, and error bar represents 95% highest density interval.

depends on the location of the decoy (Huber et al., 1982; Wedell, 1991). To simulate the varying strength, choice probability is computed with three types of decoys: Cars D, F and R (Figure 6.7a). Figure 6.7c summarizes the probabilities of choosing Cars A and B.

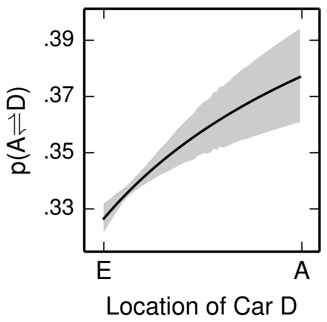
The model predicts the strongest effect with Car D. Car D is inferior to Car A in both attribute dimensions, and thus every comparison between Cars A and D favors A. In contrast, Cars R and F is inferior to A only in one dimension, and hence only half of comparison between A and the decoy (Car R or F) favors A, resulting in the weaker effect.

Also compared to Car F, Car R is further away from Car B. As a result, Car R is less frequently compared against Car B than Car F is, and hence, Car R is more frequently compared against Car A than Car F is (Figure 6.7b). As the comparisons between Cars A and R favor Car A, the difference in comparison probability leads to the difference in the probabilities of choosing Car A.

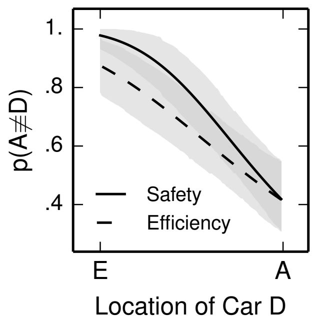
The difference between the strengths of the attraction effect with Cars R and



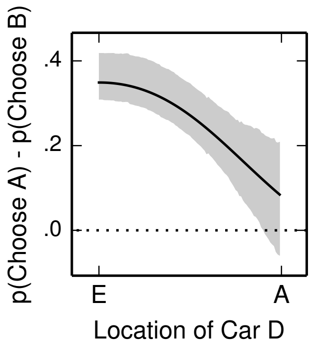
(a) Location of tested decoy



(b) Comprison probability

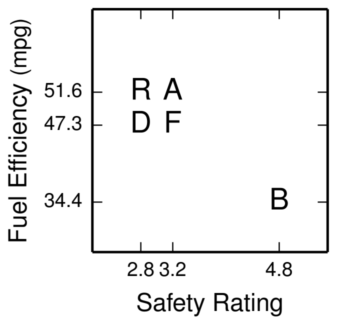


(c) Just meaningful difference

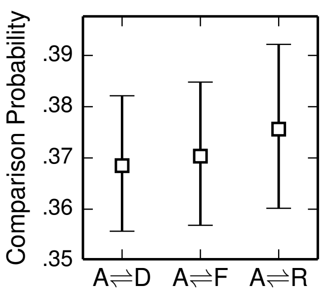


(d) A–B difference in choice probability

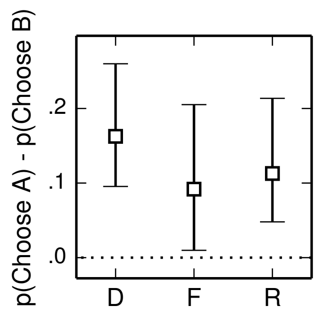
Figure 6.6: MDbS prediction of the attraction effect with various locations of Car D, from E to A (endpoint exclusive). The shaded area indicates 95% highest density intervals.



(a) Various decoys

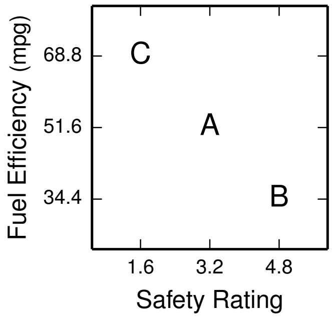


(b) Comparison probability

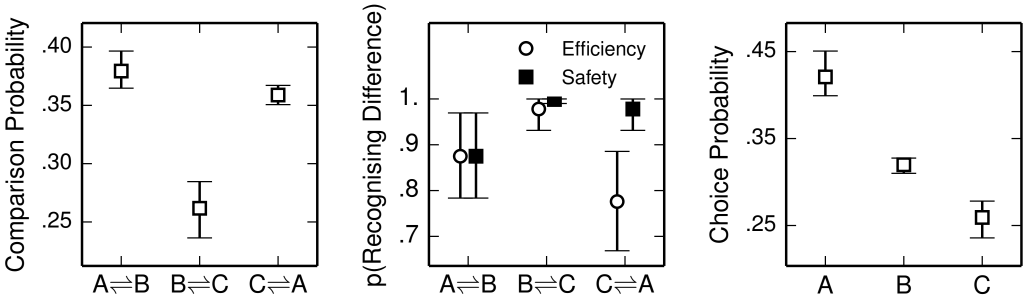


(c) A–B difference in choice probability

Figure 6.7: MDbS prediction of the attraction effect with various decoys. Error bar represents 95% highest density interval.



(a) Compromise choice set



(b) Comparison frequency (c) Just meaningful difference (d) Choice probability

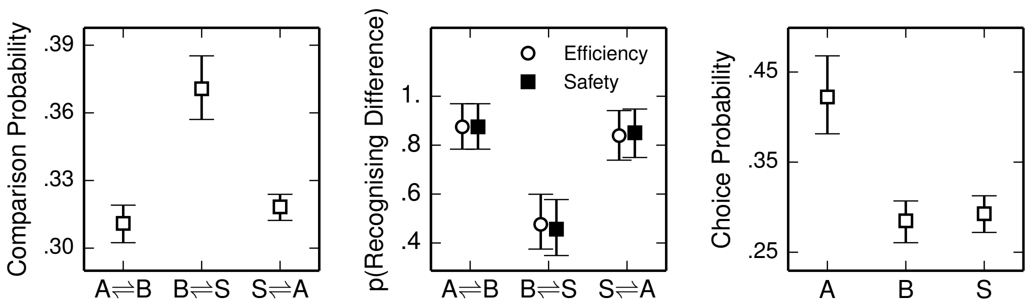
Figure 6.8: MDbS prediction of the compromise effect. Square marker shows prediction with the posterior median parameters, and error bar represents 95% highest density interval.

F is reported significant in Huber et al. (1982) but non-significant in Wedell (1991), which implies that the difference may be small. Also, the differences between the strengths with Cars R and D, and Cars D and F are non-significant in both Huber et al. (1982) and Wedell (1991). Our prediction results show the overlapping error bars.

Compromise effect

The prediction results for the compromise effect are displayed in Figure 6.8. As Car B is more similar to Car A than to Car C, the comparison between Cars A and B is more frequent than between Cars B and C. Similarly, the comparison between Cars C and A is more frequent than between Cars B and C (Figure 6.8b). As half of these frequent comparison favors Car A, Car A becomes more likely chosen with the presence of Car C (Figure 6.8d).

(a) Similarity choice set



(b) Comparison frequency (c) Just meaningful difference (d) Choice probability

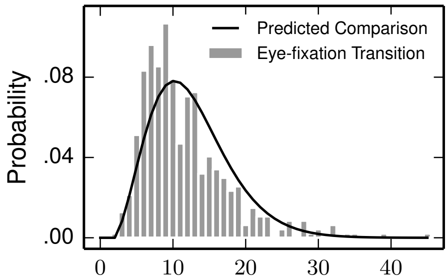
Figure 6.9: MDbS prediction of the similarity effect. Square marker shows prediction with the posterior median parameters, and error bar represents 95% highest density interval.

Similarity effect

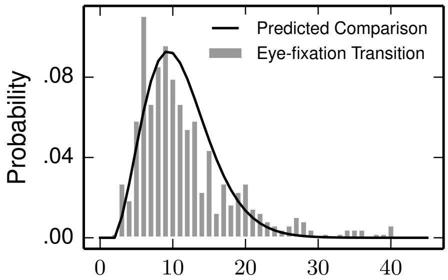
The prediction results for the similarity effect are summarized in Figure 6.9. As discussed above, Cars B and S are most often compared against each other (Figure 6.9b), but the differences between B and S are less likely recognized as meaningful (Figure 6.9c), resulting in the larger probability of A being chosen (Figure 6.9d).

6.4.2 Choice response time

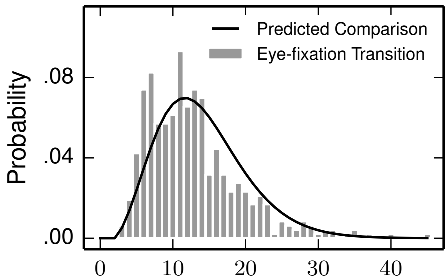
Here, we computed the number of comparisons taken in each choice set (see Appendix B.1 for the computation details). Thus we estimate the numbers of transitions using a model fit based only on choice data. The results are summarized in Figure 6.10. The solid line represents the model prediction with the posterior median parameter values. The model predicts, for instance, that there is a 7.81% probability that one alternative is chosen after exactly 10 comparisons in the attraction choice set.



(a) Attraction choice set



(b) Compromise choice set



(c) Similarity choice set

Figure 6.10: Predicted number of comparisons before choosing A, B, or T.

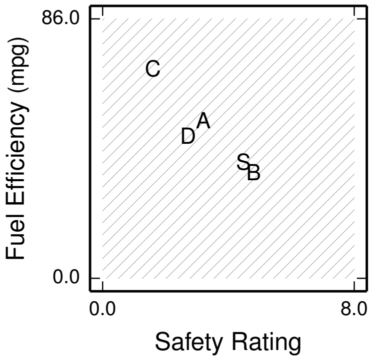
For comparison, Figure 6.10 also displays the histogram showing the number of eye-fixation transition between alternatives from Chapter 5, where participants made a choice in 10 different attraction, compromise and similarity choice sets. Although the histogram does not contain the eye-fixation data on the choice set where a choice is predicted, the distribution of the transitions are fairly well fit by the predicted number of comparisons.

The model itself and also parameter values are not designed or trained to produce this fit to the eye-fixation counts, but Figure 6.10 illustrates that the MDbS model captures the pattern of the numbers of the transitions to make a choice.

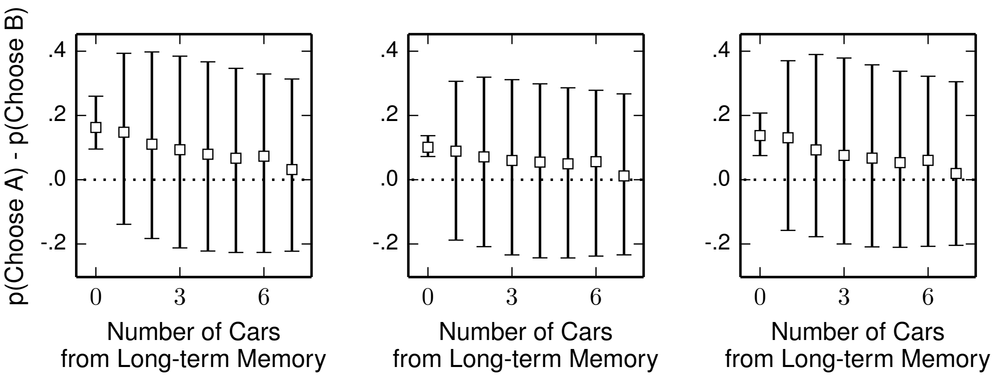
6.4.3 Effect of choice familiarity

As discussed above, the familiarity with the choice category reduces the attraction effect (Kim & Hasher, 2005) and also the compromise effect (Sheng et al., 2005). The effect of choice familiarity is explained in the MDbS model by allowing samples from long-term memory distributions of attribute values to enter working memory, in addition to the samples from the current choice set.

The cars from long-term memory in the prediction are randomly drawn from the uniform distribution, whose range is shown as the shaded area in Figure 6.11a.



(a) Memory is uniformly distributed over the shaded area



(b) Attraction effect (c) Compromise effect (d) Similarity effect

Figure 6.11: Simulated effects of choice familiarity. The square markers in Panels (b), (c), and (d) represents median predictions with the posterior median parameters. Error bars are 95% highest density interval.

These recalled cars are compared against alternatives in the choice set and remain in memory until a choice is made. Thus, to allow for comparison to Alternative M in memory, Equation 6.1 becomes:

$$\begin{aligned} p(\Psi_A\text{++}) \;=\; & p(A \rightharpoonup B) \sum_i p(\text{Attend } i)\, p(A_i \not\equiv B_i)\, 1_{A_i \geq B_i} \\ & + p(A \rightharpoonup T) \sum_i p(\text{Attend } i)\, p(A_i \not\equiv T_i)\, 1_{A_i \geq T_i} \\ & + p(A \rightharpoonup M) \sum_i p(\text{Attend } i)\, p(A_i \not\equiv M_i)\, 1_{A_i \geq M_i}. \end{aligned}$$

Because the choice probability depends on the recalled cars and recalled cars are randomly drawn from the distribution, the same parameter values can lead to different prediction on the effect of choice familiarity. Therefore, we simulated a choice 10^2 times with each of 10^4 parameters for each number of cars drawn from long-term memory. The highest density interval for these 10^6 predictions was computed for each number of cars from long-term memory and summarized in Figure 6.11. When the choice is familiar and a car is recalled, the choice probability becomes less dependent on relative advantage of the cars within the choice set. As a result, the attraction, compromise and similarity effects diminish with the number of cars recalled from long-term memory (Figures 6.11b, 6.11c, and 6.11d).

The above prediction results hold as long as the distribution of alternatives does not favor Car A over Car B or Car B over Car A. Also, as long as the median point of the distribution is at the middle point between Cars A and B, the results appears identical with other distributions (e.g., Gaussian with zero covariance).

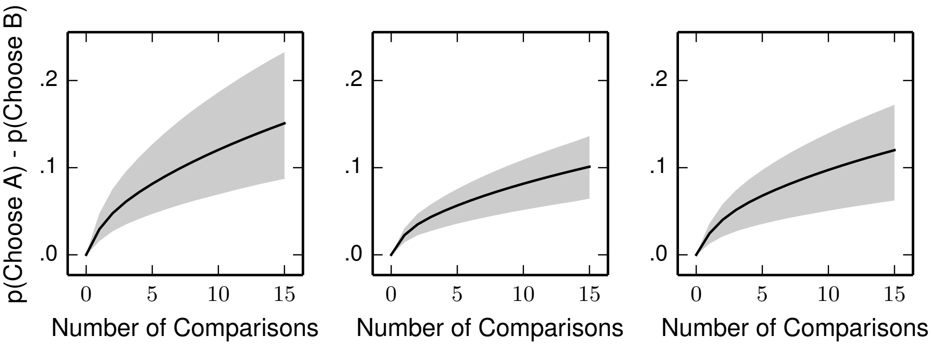
In addition, the prediction can depend on exact samples drawn from long-term memory. The simulation does not make an assumption on what sample is more or less likely to be drawn, but we assume that sampled encountered immediate past are more likely to be drawn. This assumption is further discussed when we simulate the attribute range effect below.

6.4.4 Effect of display duration

The attraction, compromise, and similarity effects are also reported to be stronger with the display duration. In the experiments reported in Pettibone (2012), three alternatives disappear from the computer screen after two, four, six or eight seconds, and participant is asked to make a choice. When the alternatives are displayed only for two or four seconds, the attraction and compromise effects are not observed or are weaker, compared to when the alternatives are displayed for eight seconds. The same effect of display duration on the similarity effect was recently reported by Trueblood et al. (in press).

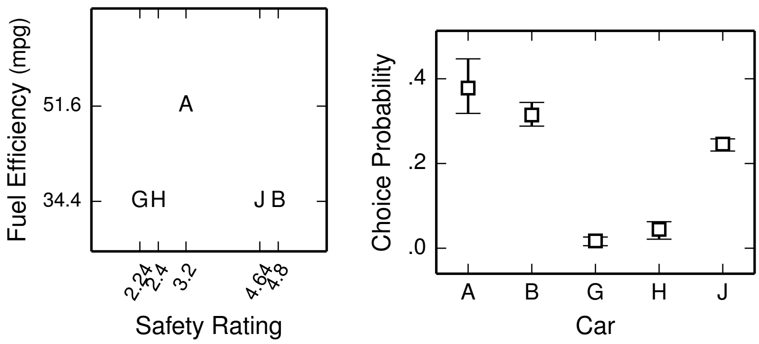
In simulating this effect of display duration, we assumed that alternatives are compared against each other only when the alternatives are displayed: rather than race to threshold, cars are compared for a fixed time after which the higher is selected. Thus we implement the external stopping rule as in decision field theory (Roe et al., 2001) (see Appendix B.2 for the computation details).

The prediction results are summarized in Figure 6.12. This figure plots the strength of the attraction, compromise, and similarity effects as a function of the number of comparisons. When no comparison is made, one alternative is randomly chosen and hence the choice probability for Cars A and B does not differ. As the number of comparisons increases, Car A becomes more likely chosen than Car B,



(a) Attraction effect (b) Compromise effect (c) Similarity effect

Figure 6.12: Simulated effects of display duration. The shaded area represents 95% highest density interval.



(a) Perceptual focus choice set (b) Predicted choice probability

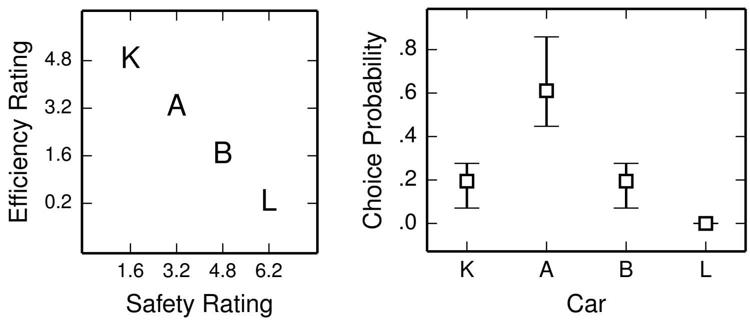
Figure 6.13: Perceptual focus effect documents the choice of Car A.

indicating the emergence of the context effects. Also in line with Pettibone (2012), the difference in the choice probability between Cars A and B is generally larger in the attraction effect than in the compromise effect at each number of comparisons.

6.4.5 Other Context Effects

The perceptual focus effect

The perceptual focus effect has been reported with the choice set in Figure 6.13a. In this choice set, four cars share the same value on the fuel efficiency, making Car A distinctive on the fuel efficiency. Hamilton, Hong, and Chernev (2007) found that Car A is more frequently chosen over Car B in the presence of Cars E, H, and J. This frequent choice of Car A is ascribed to the distinctiveness of Car A's attribute values, which facilitates an individual to focus on Car A.



(a) Attribute balance choice set (b) Predicted choice probability

Figure 6.14: Attribute balance effect documents the choice of Car A.

This perceptual focus effect is explained by the same mechanisms as in the similarity and attraction effects by the MDbS model. Although Car B dominates Car J in the safety rating and is most frequently compared against J, Car J is so similar to Car B that the difference is less likely recognized as meaningful. Meanwhile, Cars G and H are inferior to Car A in both attributes, and thus Car A is more frequently favored through the comparisons. As a result, Car A is more likely chosen than Car B (Figure 6.13b).

Attribute balance effect

The attribute balance effect is reported when two attribute dimensions are on the same scale range and unit: e.g., available cars are rated on the scale from 0 to 5 for both efficiency and safety. Under this condition, an individual tends to choose a car which has the same scores for both attributes (Chernev, 2004, 2005). An example choice set is shown in Figure 6.14a.

When the attributes are on the same scale range and unit, the dimensions are more comparable. As a result, an individual may engage in pair-wise comparison across dimensions: e.g., the efficiency rating of Car A may be compared against the safety rating of Car B. When the dimensions are collapsed, the balanced alternative becomes the compromise alternative in the choice sets used by Chernev (2004, 2005). Then, the attribute-balance effect emerges with the same mechanism as the compromise effect: Car A is most frequently compared against another Car, and as a result, Car A most frequently wins a comparison.

Thus to simulate this attribute balance effect, we assumed that the comparison frequency depends on within-dimension and across-dimension similarity. Thus, the comparison frequency between Cars A and B, for example, is proportional to the following:

$$\exp\left(-\alpha\ \left(\sum_i\sum_j \Delta_{A_i,B_j}\right)^{1/\gamma}\right).$$

Also, the four cars are compared across the attribute dimensions. For example, efficiency rating of Car A can be compared against safety rating of Car B. The prediction results are summarized in Figure 6.14b. This figure shows that the predicted choice probability is the highest for Car A, if the attribute dimensions are collapsed.

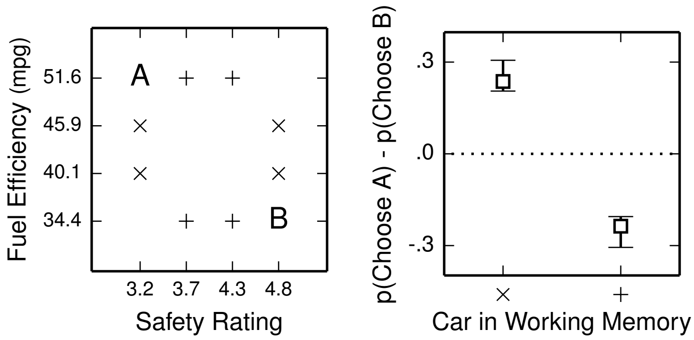
Consistent with the assumption of the collapsed dimensions, Chernev (2004) reports that when an individual is primed to examine alternatives attribute by attribute and also not to collapse the dimensions, their choices do not show the attribute-balance effect. In addition, when the attribute dimensions are not collapseable (e.g., different units), the attribute-balance effect is not observed (Chernev, 2004).

Attribute range effect

The attribute range effect (Mellers & Cooke, 1994; Cooke & Mellers, 1998) documents that how attractive an individual finds one attribute value depends on what values this individual previously saw in other choice sets. After seeing the fuel efficiency of 68.8mpg in one choice set, for example, an individual finds the fuel efficiency of 51.6mpg in another choice set less attractive. After seeing the fuel efficiency of 34.4mpg, however, an individual finds 51.6mpg more attractive. A similar effect is also reported in a risky choice (Ungemach et al., 2011).

This attribute range effect is explained through a comparison against alternatives in working memory in the MDbS model. After an individual saw the fuel efficiency of 34.4mpg in one choice set, this fuel efficiency is likely to remain on his or her working memory and to be compared against a fuel efficiency in the next choice set. If this individual recalls 34.4mpg, the memory comparison favors 51.6mpg, resulting in higher attractiveness of 51.5mpg.

To simulate this effect, we computed the probability of choosing Cars A and B assuming that $\times$ or $+$ in Figure 6.15a are in working memory. Cars marked as $\times$ enhance the advantage of Car A on the fuel efficiency, resulting in the higher choice probability for Car A than Car B (Figure 6.15b). When Cars marked as $+$ are in the working memory, instead, the advantage of Car B on the safety rating is enhanced, resulting in the higher choice probability for Car B than Car A (Figure 6.15b).



(a) Attribute range choica set (b) Choice probability difference

Figure 6.15: Attribute range effect.

Context effects in risky and intertemporal choice

When an individual is familiar with the attribute dimensions, the distributions of attribute values in this individual's long-term memory is likely to reflect the environmental distributions. With this assumption, the decision by sampling is able to explain the shapes of utility/value, probability weighting, and temporal discounting functions (Stewart et al., 2006), and also key phenomena in risky choice (e.g., the common ratio effect) (Stewart & Simpson, 2008; Stewart, 2009). Recently, Stewart et al. (in press) demonstrate that the shapes of utility/value and probability weighting functions can be manipulated by changing the environmental distribution of gains, losses, probabilities, and delays, supporting the predictions from the decision by sampling model.

6.5 Model Comparison

This section compares the performance of the MDbS model against that of multi-alternative decision field theory (MDFT; Roe et al., 2001) in predicting the attraction, compromise and similarity effects. MDFT is among the most extensively studied models of multi-alternative choice. Berkowitsch et al. (in press), for example, compare MDFT against probit models and demonstrate the advantage of MDFT. Also, Trueblood et al. (in press) compare MDFT against multi-attribute linear ballistic accumulator model.

The MDFT model is an extension of decision field theory (Busemeyer & Townsend, 1993) to explain the three context effects under the single frame work. As in the MDbS model, the MDFT implements dynamic preference development, where a choice is reached through a series of evaluations and comparisons. Unlike the

MDbS model, however, an alternative is evaluated in relation to the mean attribute values of the other alternatives, and the developing preferences inhibit each other, where the strength of the inhibition depends on the distance between alternatives. The relative evaluation and distance-based inhibition is the key component of MDFT to explain the context effects (at the posterior median, the parameter values are $\phi_1 = 2.02$, $\phi_2 = 1.01$, $\xi = 15.39$, $\sigma^2 = 0.02$, and a threshold is 1.89).

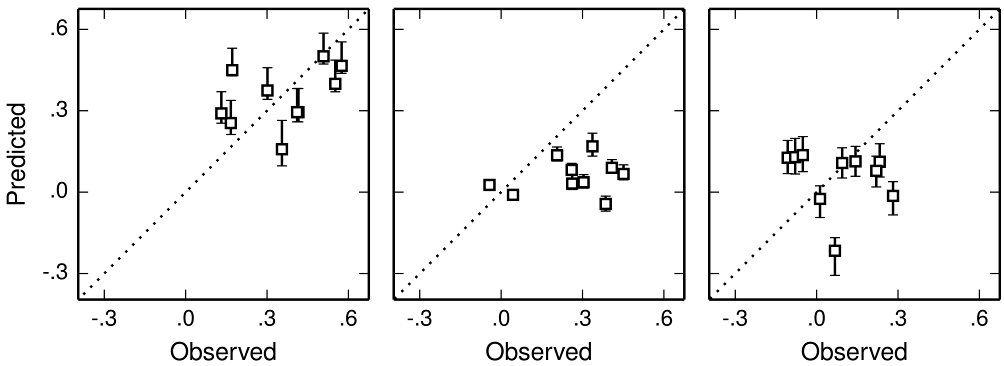
The MDFT model has been examined by assuming that a choice is reached after a fixed amount of time (an external stopping rule) (e.g., Trueblood et al., in press) or after preference is stabilized (a convergence stopping rule) (e.g., Berkowitsch et al., in press), rather than by assuming that a choice is reached after preference for one alternative reaches a threshold (an internal stopping rule). In Chapter 5, the timing of choice was not controlled: participants were allowed to spend as much time as necessary to make a choice. Therefore, the experiment design reflects the use of the convergence stopping rule or the internal stopping rule.

The convergence stopping rule, however, is not applicable to the models without inhibition, including the MDbS model. For the purpose of the model comparison, we consider it fair to use the same stopping rule for both MDbS and MDFT models, and thus compared the models with the internal stopping rule.

The MDFT model with the internal stopping rule, however, does not have an analytic solution to choice probability, and the choice probability has been derived by simulating choices (e.g, Hotaling, Busemeyer, & Li, 2010). When choice probability is derived by simulations, the same parameter value can lead to slightly different predictions. To reduce this uncertainty, we simulated 20,000 choices to derive prediction and manually seeded the random number generator, so that the same parameter values lead to the reasonably accurate, identical predictions.

Lastly unlike the MDbS model, the MDFT model is scale-variant: when the attribute values are on the different scale, one attribute dimension could disproportionally drive preference development. This scale-variance make it impossible to apply the same parameter values across the different choice sets with different units. Thus, it is necessary to normalize the attribute values for the MDFT model. Following the personal communication with Prof. Jerome R. Busemeyer (2013), we normalized the values to range from 0 to 1, using the maximum and the minimum value on each attribute dimensions across choice sets in the entire experiment (see Appendix B.3 for the more details).

Then we sampled posterior parameter values for the MDFT model using the same method as we used for the MDbS model as described earlier: We pooled the choice response data from Chapter 5, and used a non-informative, improper



(a) Attraction Choice Set. (b) Compromise Choice Set. (c) Similarity Choice Set.

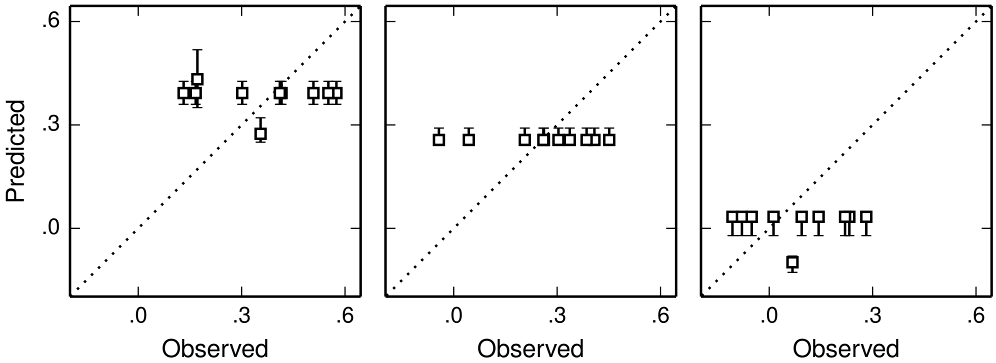
Figure 6.16: MDbS posterior predictives for $p(\text{Choose A})-p(\text{Choose B})$. The square markers show the prediction with the posterior median parameters ($\alpha = 0.056$, $\beta_0 = 15.01$, $\beta_1 = -14.65$ $\gamma = 0.36$, and $\lambda = 3$), and the error bars show the 95% highest density interval. The cross represents overall mean.

distribution as the prior. This prior distribution, however, covers only the area where preference can be stable: i.e., the eiganvalue of the inhibition matrix is equal to or less than one in magnitude (Roe et al., 2001). Outside this stable area, the prior distribution is set to 0. Also due to the computational expense, we sampled the 10^5 parameter values, instead of 10^6, discarded the first half, and thinned at an equal interval to retain 10^4 samples.

6.5.1 Posterior predictive distributions

Figures 6.16 and 6.17 show the model predictions on the choice sets we used to draw the posterior parameter values. Each marker represents a choice product (e.g., cars, laptops, and TV sets). Chapter 5 had two versions for each choice product: one where the context favors A, and the other where the context favors B. Alternatives in the latter version are recoded so that the context favors A, and the predictions for the two versions are mean-averaged.

While the MDbS model predicts the overall attraction effect reasonably well (Figure 6.16a), the model underpredicts the overall strength of the compromise effect (Figure 6.16b) and does not provide an accurate variability in the strength of the similarity effect between choice products (Figure 6.16c). This observation confirms that the compromise and similarity effects are on a fine balance in the MDbS components. When the similarity between alternatives has a stronger influence on the comparison probability, the compromise effect becomes stronger but the similarity effect becomes weaker. Also when the just-meaningful-different threshold is higher,



(a) Attraction Choice Set. (b) Compromise Choice Set. (c) Similarity Choice Set.

Figure 6.17: MDFT posterior predictives for $p(\text{Choose A}) - p(\text{Choose B})$. The square markers show the prediction with the posterior median parameters ($\phi_1 = 2.02$, $\phi_2 = 1.01$, $\xi = 15.39$, $\sigma^2 = 0.02$, and a threshold is 1.89), and the error bars show the 95% highest density interval. The cross represents overall mean.

the compromise effect becomes weaker but the similarity effect becomes stronger. For the MDbS model, this balance resulted in the underprediction of the strength of the compromise effect.

The same, more subtle balance leads the MDFT model to underpredict the strength of the similarity effect (Figure 6.17c). Also while the MDFT model captures the attraction, compromise and similarity effects on average, the model is not able to predict the variability in the strength of context effects between choice problems, primarily because of the normalization (Figure 6.17).

6.5.2 Predictive accuracy

Here, we compare performance of the MDbS and MDFT models. In comparing computational cognitive models, previous research has often used Baysian information criteria (BIC; Schwarz, 1978) or the odds of the model likelihoods. BIC approximates the model likelihood, when the number of observation in data is large, and when the distribution of the prediction errors follows an exponential family (e.g., Gaussian, binomial and multinomial). While our application of the MDbS and MDFT models assumes the multinomially distributed error, it is unrealistic to use BIC here, as the model likelihood is not available when the prior distribution is improper (Gelman et al., 2013). More generally, the value of the model likelihood depends on the exact specification of the prior. For example, the model likelihood, computed with the prior distribution for one parameter being the uniform distribution between -100 and 100, can be lowered by the factor of 10, when the prior for the

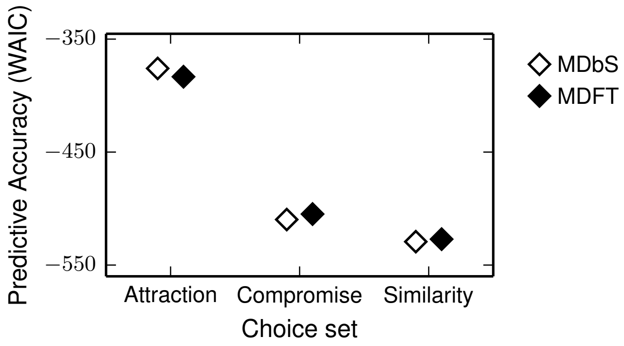


Figure 6.18: Predictive accuracy of the models.

parameter is the uniform between -1000 and 1000. Moreover, the model likelihood depends on the model parameterization: the MDbS model can have larger or smaller model likelihood by employing the slightly altered function for the just-meaningful-differences while keeping the model predictions identical. These properties led us to conclude that the model likelihood may not be particularly useful in comparing distinctively different models, such as the MDbS and MDFT models (see also Gelman & Shalizi, 2013).

Instead, we compared the models in their predictive accuracy and assessed the out-of-sample prediction error. While a natural way to assess this error is cross-validation, the cross-validation can be computationally very expensive. Thus for practical reasons, several approximation methods have been proposed (for review, see Gelman, Hwang, & Vehtari, in press), and we used one of the most accurate method, Widely Applicable Information Criteria (WAIC; Watanabe, 2010). In its computational specification, WAIC penalizes the predictive accuracy with the flexibility of the prediction. Thus the model is penalized when the posterior predictive distribution is off the observed choice proportion and/or has a large variance (see Appendix B.4 for the computational detail).

WAIC for the MDbS and MDFT models are displayed in Figure 6.18. The larger value of WAIC indicates the better performance. As we have the same number of observations for the attraction, compromise, and similarity choice sets, WAIC is comparable across the three choice sets. Across the choice sets, both models show the same pattern of predictive accuracy: both MDbS and MDFT models make more accurate prediction for the attraction choice set than for the compromise or similarity choice set. This reduced accuracy in the compromise and similarity choice sets quantitatively confirms that the difficult balance the models are making in predicting the strength of the compromise and similarity effects.

The advantage of the MDbS model in predicting the attraction effect outweights the smaller advantage of the MDFT in predicting the compromise and similarity effects. As a result, the MDbS model (WAIC = -1411.28) outperforms the MDFT model (WAIC = -1413.10) for overall accuracy. Thus, while both the MDbS model and the MDFT predicts the attraction effect better than the compromise and similarity effects, the MDbS model provides a better prediction of the attraction effect. This advantage results in the overall better predictive accuracy of the MDbS model.

6.6 Relations to Other Models

We proposed a new model of decision-making: the MDbS model. This new model has the three components: 1) an alternative is evaluated through a series of pairwise comparisons on a single attribute value, 2) similar alternatives are compared more often, and 3) relatively small differences in attribute values are ignored. We have demonstrated that these components explain the attraction, compromise, and similarity effects. These components of the MDbS model, however, have been implemented in existing models of decision making. In this section, we address how the MDbS model relates to other models.

Recently, Simonson et al. (2013) proposed that choices are based on the comparisons which are task-acceptable and easy to make. According to Simonson et al. (2013), the task-acceptability depends on whether the comparison results are informative in judging which alternative is better. For example, when choosing between Cars A and B, a comparison between Cars A and C is not acceptable. This is because a comparison between Cars A and C does not justify a choice of Car A over B or Car B over A. Therefore, alternatives are compared only within a choice set, and inevitably, this account does not readily explain the effects of choice familiarity and the attribute-range effect. The other component of Simonson et al. (2013)'s proposal, the ease of comparison, depends on a number of factors, including computational ease and saliency of alternatives. With this regard, we propose that similarity between alternatives also determines probability of comparisons.

However, Simonson et al. (2013) does not specify how alternatives are evaluated in the comparisons. In the MDbS model, the comparison is insensitive to the magnitude of the differences, as long as the difference is judged meaningful. The magnitude-insensitive comparison is implemented in the model proposed by de Clippel and Eliaz (2012). In de Clippel and Eliaz (2012), each alternative is ranked on each attribute dimension, and an individual chooses the alternative whose minimum ranking is highest. Thus, the procedure to rank alternatives is unbiased: unlike the MDbS model, the rank is not influenced by similarity between alternatives.

Influence of similarity on decision making has also been employed in models of risky choice (e.g. Rubinstein, 1989; Leland, 1994; Buschena & Zilberman, 1999). For example, Buschena and Atwood (2011) argue that an individual employs different decision strategies depending on the similarity between alternatives. While in the MDbS model similar alternatives are compared in the same manner as dissimilar alternatives, the similarity between alternatives determines the probability of comparison and also the probability of recognizing the difference as meaningful. These influence of similarity on choice processes might appear as if an individual is engaged in different choice processes depending on the similarity.

Also among the models of risky choice, priority heuristic (Brandstatter et al., 2006) implements the just meaningful difference. This heuristic predicts that an individual chooses the alternative if the alternative excels another by 10%. Brandstatter et al. (2006) argue that this 10% threshold is fixed. In contrast, the threshold for the just meaningful difference is probabilistic in the MDbS model. The probabilistic threshold has been implemented in models to explain how a change in prices influences consumer behavior (e.g., Han, Gupta, & Lehmann, 2001) and also to explain choices of transportation (e.g., Cantillo, Heydecker, & Ortúzar, 2006; Cantillo & Ortúzar, 2006).

The MDbS model is perhaps most similar to the exemplar-based random walk (EBRW) model of categorization (Nosofsky & Palmeri, 1997). According the EBRW model, an individual categorizes a target object through a series of pair-wise comparisons. In each comparison, an object is recalled from memory, and as in the MDbS model, the target object is more likely compared against a similar object, and the comparison results are accumulated over time. Also, similar to the MDbS model, when this accumulation reaches a threshold, the individual makes a choice in the EBRW model. Unlike the MDbS model however, the choice in the EBRW model is on what category the target object belongs to, and the individual accumulates counts for categories a compared object favors. In the MDbS model, the individual accumulates counts for alternatives a comparison favors.

The three components of the MDbS model have been employed in various models. Our contribution to the exiting literature is to show that combination of the components explains the context effects.

6.7 Limitations

The multi-alternative decision by sampling model, we propose here, concerns a choice in a static choice set, which excludes a non-static choice set. In a non-static choice set, an alternative becomes unavailable or available at an arbitrary time point. When an alternative is announced unavailable, this announcement often changes the preference order (Worchel, Lee, & Adewole, 1975; see Verhallen & Robben, 1995, for review). Boland, Brucks, and Nielsen (2012), for example, asked participants to choose the most and second most preferred alternatives in a choice set. When the most preferred alternative is announced unavailable, the second most preferred alternative does not become the most preferred among the available alternative. Instead, another alternative, which is similar to the unavailable alternative, becomes the most preferred among the available alternatives. This effect of alternative unavailability is also called the phantom decoy effect (Highhouse, 1996). When an alternative becomes unavailable, a choice process appears to change (e.g., Walter & Festinger, 1964). While the nature of this process change is an interesting topic, our intention in this article is focused on a choice in a static choice set.

6.8 Conclusion

In this chapter, we developed the decision by sampling account of the attraction, compromise, similarity, and other context effects. The preference for an alternative is developed through a series of pair-wise comparisons on single attribute dimension. In the comparisons, similar alternatives are more often compared, but a relatively small difference is considered to be not meaningful. We have demonstrated that these three components explain a range of empirical findings concerning the context effects.

Chapter 7

Conclusion

This thesis investigated how an individual evaluates an alternative and how he or she makes a choice. Choices have been extensively studied and modeled with two classes of models: static and dynamic models. Traditionally, the class of static models has been developed to understand risky choice. In risky choice, an alternative is associated with probabilistic pay-offs, and models of risky choice have attracted much attention in psychology, and in experimental and behavioral economics. As I discussed in the introduction, these models have typically been examined using carefully selected sets of choice alternatives (e.g., Allais's paradox to compare prospect theory against expected utility theory).

A model performance however, can depend on which alternatives are being considered. A model, which outperforms another model in one set of alternatives, may underperform in another set of alternatives. To evaluate models with a wide range of alternatives, Chapter 2 developed a non-parametric method for estimating the utility map and comparing the models. The estimated maps are compared against three of the most well-known static models: expected utility theory, cumulative prospect theory, and the transfer of attention exchange model. The results show that cumulative prospect theory and the transfer of attention exchange model fits better to choices than expected utility theory, indicating an advantages of cumulative prospect theory and transfer of attention exchange model over expected utility theory.

However, the results in Chapter 2 also show that utility predicted from the models deviates from the estimated map in some regions of the probability triangle. The deviation implies that a new model, if it better captures patterns of estimated map, could outperform the existing models. Utility, however, may be unstable, as the results from the following chapters indicate.

Chapter 3 examined effects of set-sizes and information presentation formats in risky choice. Previous research on risky choice, including the study reported in Chapter 2, has been predominantly based on studies of choices between two alternatives, with the findings often generalized to environments with more than two alternatives. One prominent claim of this research is that choices differ with respect to risk when alternatives are described — the description paradigm — as opposed to experienced — the experience paradigm: Individuals appear to make choices as if they over-weight small probabilities in the description paradigm, but under-weight the same probabilities in the experience paradigm.

Chapter 3 demonstrated that the under-weighting in the experience paradigm is sensitive to set-sizes in the gain domain. Two experiments show that with a growth in set-sizes, choices systematically favor risky alternatives in the experience paradigm, appearing as if small probabilities are over-weighted. This risk-amplification is due to the statistical structure of pay-offs, where in a large set, at least one risky alternative is more likely to deliver a pay-off at a much higher frequency than its underlying probability. Chapter 3 simulated several static models to further demonstrate that the difference between the frequency of pay-off and its true probability drives the risk-amplification, regardless of choice and search strategies employed.

Effects of set-sizes on risky choice are further explored in Chapter 4. In particular, Chapter 4 examined information overload. A large set was previously reported to overload an individual's cognition — information overload — often resulting in choice deferral. More recent studies, however, indicate that information overload is more tightly associated with the number of attributes to describe alternatives than with set-sizes. Thus, Chapter 4 assessed impacts of the number of possible pay-offs on choice deferral: participants were endowed with money and then given the opportunity to search through varying numbers of alternatives and to either purchase a pay-off from an alternative or to defer a choice and keep the money. The alternatives were presented in the description or experience paradigm.

The results in Chapter 4 show that in both description and experience paradigms, individuals are more likely to purchase an alternative when the set size is larger. Also in the description paradigm, choice is more frequently deferred when an alternative is associated with many possible pay-offs than with few possible pay-offs. The results indicate that a growth in set-sizes leads to thin-search — search for less information per alternative — and that an increment in the number of attributes promotes non-compensatory strategy use. Thin-search often results in a systematic bias in evaluation of alternatives. This bias explains choice deferral.

These results from Chapters 3 and 4 indicate that utility may depend on information presentation formats and also the number of possible pay-offs. In the experience paradigm, evaluation of an alternative is often influenced by the statistical structure of pay-offs, and with a growth in set-sizes, this influence eventually overrides subtle differences between static models. Also, the results from Chapter 4 indicate that one model may explain choices the best when an alternative is associated with few possible pay-offs, but that the same model may be outperformed by another model when the number of possible pay-offs is increased.

These results imply an alternative way to consider static models. As discussed in the introduction, static models are often developed with an aim to produce single model that explains as many choice phenomena (e.g., violation of coalescing) as possible. The results presented in Chapter 4, however, indicate that an individual may adapt a different strategy depending on the structures of alternatives. Thus, the results imply that existing static models may be too specific in explaining choices in one context, and exiting models may lack generalizability to explain choices across contexts and structures of alternatives.

The instability of utility, however, has been captured by dynamic models, and with this regard, dynamic models might be able to provide a better explanation of choices than static models. Unlike static models, a dynamic model assumes that an individual iteratively and stochastically develops preferences for alternatives. This stochastic nature of preference development can capture the instability of utility. While dynamic models can be compared on their accuracy to predict choices, dynamic models can also be evaluated on assumed processes of preference development. The assumptions on choice processes are tested in Chapter 5.

In particular, Chapter 5 examined eye-movements recorded during a series of three alternative choice. In three alternative choice, the attraction, compromise, and similarity effects demonstrate instability of utility: the utility of an alternative appears dependent on the other alternatives in a set. Thus, these effects are considered to suggest that utility is realized through the comparison of alternatives. Chapter 5 investigated exactly how alternatives are compared against each other. The results indicate that a series of comparisons is made in each choice, with a pair of alternatives compared on a single attribute dimension in each comparison. Then, Chapter 5 concluded that psychological models of choice should be based on the single-attribute pair-wise comparisons, as these comparisons are not assumed in two of the dynamic models reviewed in the introduction: decision field theory (Busemeyer & Townsend, 1993; Roe et al., 2001) and comparison grouping model (Tsuzuki & Guo, 2004).

Following the conclusion from Chapter 5, Chapter 6 proposes the multi-alternative decision by sampling model. This new dynamic model explains the attraction, compromise, and similarity effects with three components: an alternative is evaluated through a series of pair-wise comparisons on single attribute values, similar alternatives are compared more often, and relatively small differences in attribute values are ignored. The first two components are supported by eye-movement data in Chapter 5, and the third component is based on well-established empirical findings. Thus, the proposed model is tightly grounded on choice processes implicated by the eye-movement data and previous empirical findings.

The model evaluation reported in Chapter 6 indicates that the proposed model explains the attraction, compromise, and similarity effects equally well to or slightly better than decision field theory. This result demonstrates again that dynamic models, including the proposed model, are more difficult to differentiate in their ability to explain choices than in their assumptions on choice processes. The proposed model, however, can reproduce empirical findings related to choice processes and also predicts various other context effects, demonstrating that the proposed model accurately reflects choice processes.

Taken together, the studies reported in this thesis imply an advantage of dynamic models over static models, as a core assumption of static models — presence of stable utility — is not supported. As reviewed in Chapter 1, static models explain choices with the utility maximization principle. With this principle, various static models have been proposed to provide a descriptive account of utility, assuming that utility is stable across various contexts determined by information presentation formats and relationships between alternatives within a set. As a result, static models typically do not account for instability of utility. In contrast, dynamic models could incorporate instability of utility, and in Chapter 6, the new, dynamic model was demonstrated with its ability to produce various empirical findings.

The two classes of models, however, were not directly compared against each other in this thesis: Static models are discussed in Chapters 2, 3, and 4, and dynamic models are discussed in Chapters 5 and 6. While Chapter 5 successfully employed eye-movement data to evaluate choice processes assumed by dynamic models, eye-movement data were not used to test assumptions of static models. Thus, the studies reported here evaluated static and dynamic models with different methods, and hence, this thesis does not provide a direct, empirical support for an advantage of dynamic models. Thus, the advantage of dynamic models is speculative as it is.

The studies reported here, however, indicate that models, either static or dynamic, can be difficult to differentiate in their ability to explain choices (see also

Birnbaum, 2011), and that it is often required to evaluate models with additional data to test their assumptions.

Possible advantage of dynamic models over static models, however, does not invalidate static models altogether. Static models have been used to understand rationality of choices (Oaksford & Chater, 2009), and thus, static models can provide a guidance for rational choices. Also, while dynamic models can explain instability of utility due to various contexts, dynamic models may not be able to explain instability of utility due to other psychological factors, such as emotion, fatigue, and stress.

To summarize, this thesis evaluated various models on their ability to explain choices and their assumptions of choice processes, and based on empirical findings on choice processes, the new model was proposed. I would like to conclude this thesis with a hope that the studies reported here will be leveraged in future research to further advance scientific understanding of choices.

References

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science*, *57*, 1879–1895.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole Americaine. *Econometrica*, *21*, 503–546.

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (in press). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*, 23–36. (Original work published 1738)

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*, 522–543.

Bird, G. D., Lauwereyns, J., & Crawford, M. T. (2012). The role of eye movements in decision making and the prospect of exposure effects. *Vision Research*, *60*, 16–21.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, *10*, 399–407.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501.

Birnbaum, M. H. (2011). Testing theories of risky decision making via critical tests. *Frontiers in Psychology*, *2*, 315.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, *71*, 161–194.

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: why both are important for analyzing decisions. *Journal of Applied Econometrics*, *986*, 963–986.

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1485), 1655–70.

Boland, W. A., Brucks, M., & Nielsen, J. H. (2012). The attribute carryover effect: What the "runner-up" option tells us about consumer choice processes. *Journal of Consumer Research*, *38*, 872–885.

Bonaccio, S., & Reeve, C. L. (2006). Consideration of preference shifts due to relative attribute variability. *Organizational Behavior and Human Decision Processes*, *101*, 200–214.

Brandstatter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409-432.

Bröder, A. (2000). Assessing the empirical validity of the "Take-The-Best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1332–1346.

Browne, G. J., & Pitts, M. G. (2004). Stopping rule use during information search in design problems. *Organizational Behavior and Human Decision Processes*, *95*, 208–224.

Buschena, D. E., & Atwood, J. A. (2011). Evaluation of similarity models for expected utility violations. *Journal of Econometrics*, *162*, 105–113.

Buschena, D. E., & Zilberman, D. (1999). Testing the effects of similarity on risky choice: Implications for violations of expected utility. *Theory and Decision*, *46*, 253–280.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, *2*, 61–104.

Cantillo, V., Heydecker, B., & Ortúzar, J. D. D. (2006). A discrete choice model incorporating thresholds for perception in attribute values. *Transportation Research Part B-Methodological*, *40*, 807–825.

Cantillo, V., & Ortúzar, J. D. D. (2006). Implications of thresholds in discrete choice modelling. *Transport Reviews*, *26*, 667-691.

Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, *31*, 131–145.

Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, *30*, 170–183.

Chernev, A. (2004). Extremeness aversion and attribute-balance effects in choice. *Journal of Consumer Research*, *31*, 249–263.

Chernev, A. (2005). Context effects without a context: attribute balance as a reason for choice. *Journal of Consumer Research*, *32*, 213–223.

Cook, G. J. (1993). An empirical investigation of information search strategies with implications for decision support system design. *Decision Sciences*, *24*, 683–699.

Cooke, A. D. J., & Mellers, B. A. (1998). Multiattribute judgment: Attribute spacing influences single attributes. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 496–504.

de Clippel, G., & Eliaz, K. (2012). Reason-based choice: A bargaining rationale for the attraction and compromise effects. *Theoretical Economics*, *7*, 125–162.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research*, *24*, 215–231.

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS and related disciplines. *The Information Society*, *20*, 325–344.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., . . . Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*, 15–47.

Ert, E., & Erev, I. (2007). Replicated alternatives and the role of confusion, chasing, and regret in decisions from experience. *Journal of Behaviora Decision Making*, *20*, 305–322.

Flinn, P. A., & McManus, G. M. (1961). Monte Carlo calculation of the order-disorder transformation in the body-centered cubic lattice. *Physical Review*, *124*, 54–59.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*, 75–117.

Fox, C. R., & Hadar, L. (2006). Decisions from experience = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*, 159–161.

Futrell, C. M., & Varadarajan, P. R. (1985). Marketing executives' perceptions of equitable salary increases. *Industrial Marketing Management*, *14*, 59–67.

Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social sampling explains apparent biases in judgments of social environments. *Psychological Science*, *23*, 1515–

1523.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). London: CRC Press.

Gelman, A., Hwang, J., & Vehtari, A. (in press). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*.

Gelman, A., & Shalizi, C. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*, 8–18.

Glaholt, M. G., & Reingold, E. M. (2009). The time course of gaze bias in visual decision tasks. *Visual Cognition*, *17*, 1228–1243.

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, *4*, 125–146.

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*, 71–98.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review*, *118*, 523–551.

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science*, *18*, 240–246.

Greifeneder, R., Scheibehenne, B., & Kleber, N. (2010). Less may be more when choosing is difficult: Choice complexity and too much choice. *Acta Psychologica*, *133*, 45–50.

Hamilton, R., Hong, J., & Chernev, A. (2007). Perceptual focus effects in choice. *Journal of Consumer Research*, *34*, 187–199.

Han, S., Gupta, S., & Lehmann, D. R. (2001). Consumer price sensitivity and price thresholds. *Journal of Retailing*, *77*, 435–456.

Harless, D. (1992). Predictions about indifference curves inside the unit triangle. *Journal of Economic Behavior and Organization*, *18*, 391–414.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: the role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*, 493–518.

Helgeson, J. G., & Ursic, M. L. (1993). Information load, cost/benefit assessment and decision strategy variability. *Journal of the Academy of Marketing Science*, *21*, 13–20.

Heneman III, H. G., & Ellis, R. A. (1982). Behavioral and industrial relations perspectives on compensation: Contributed papers. *Labor Law Journal*, *33*, 533–538.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534–539.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*, 517–23.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*, 225–237.

Highhouse, S. (1996). Context-dependent selection: The effects of decoy and phantom job candidates. *Organizational Behavior and Human Decision Processes*, *65*, 68–76.

Hilbig, B. E., & Gloeckner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, *138*, 390–396.

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science*, *21*, 1787–1792.

Hills, T. T., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review*, *20*, 1023-1031.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1351–1381.

Hotaling, J. M., Busemeyer, J. R., & Li, J. (2010). Theoretical developments in decision field theory: Comment on Tsetsos, Usher, and Chater (2010). *Psychological Review*, *117*, 1294–1298.

Huber, J., Payne, J., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98.

Iyengar, S. S., Huberman, G., & Jiang, W. (2004). How much choice is too much: Determinants of individual contributions in 401K retirement plans. In O. S. Mitchell & S. Utkus (Eds.), *Pension design and structure: New lessons from behavioral finance* (pp. 83–95). Oxford, England: Oxford University Press.

Iyengar, S. S., & Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, *94*, 530–539.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 143–150.

Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, *11*, 63–69.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Similarity, kernels, and the triangle inequality. *Journal of Mathematical Psychology*, *52*, 297–303.

Kahneman, D., & Tversky, A. (1979). Prospect theory - Analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kalwani, M. U., & Yim, C. K. (1992). Consumer price and promotion expectations: an experimental study. *Journal of Marketing Research*, *29*, 90–100.

Kim, S., & Hasher, L. (2005). The attraction effect in decision making: Superior performance by older adults. *Quarterly Journal of Experimental Psychology*, *58A*, 120–1333.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*.

Kullback, S., & Leibler, P. C. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86.

Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*(6604), 74–77.

Lajarraga, Y., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*, 334–342.

Leland, J. W. (1994). Generalized similarity judgment: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, *9*, 151–172.

Loomes, G. (2010). Modeling choice and valuation in decision experiments. *Psychological Review*, *117*, 902–924.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, *65*, 581–598.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York: John Wiley & Sons, Inc.

Lussier, D. A., & Olshavsky, R. W. (1979). Task complexity and contingent processing in brand choice. *Journal of Consumer Research*, *6*, 154–165.

Malhotra, N. K. (1982). Information load and consumer decision making. *Journal of Consumer Research*, *8*, 419–431.

Mellers, B. A., & Cooke, A. D. J. (1994). Trade-offs depend on attribute range. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1055–1967.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087-1092.

Meyer, D. E., & Irwin, D. E. (1981). On the time course of rapid information-processing. *Bulletin of the Psychonomic Society*, *18*, 68.

Myung, I. J., & Pitt, M. A. (1997). Applying Occams razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, *4*, 79-95.

Neal, R. (1993). *Probabilistic inference using Markov chain Monte Carlo methods* (Tech. Rep. No. CRG-TR-93-1). Department of Compute Science, University of Toronto.

Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review*, *14*, 1133–1139.

Noguchi, T., & Hills, T. T. (under review). Set-size induced risk-amplification: Experience-based decisions in large set-sizes favour riskier alternatives.

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*, 44–56.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, *32*, 69–84.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Bahvior and Human Performance*, *16*, 366–387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. *Judgment and Decision Making*, *7*, 513–523.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–527.

Rambo, W. W., & Pinto, J. N. (1989). Employees' perception of pay increases. *Journal of Occupational Psychology*, *62*, 135–145.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory : A dynamic connectionist model of decision making. *Psychological*

Review, *108*, 370–392.

Rosenthal, J. S. (2011). Optimal proposal distributions and adaptive mcmc. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 93–112). Florida: CRC Press.

Rubinstein, A. (1989). Similarity and decision making under risk: Is there a utility theory resolution to the allais paradox? *Journal of Economic Theory*, *46*, 145–153.

Russo, J. E. (1974). More information is better: A reevaluation of Jacoby, Speller and Kohn. *Journal of Consumer Research*, *1*, 68–72.

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 676–696.

Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, *21*, 274–290.

Russo, J. E., & Rosen, L. D. (1975). An eye fixation analysis of multialternative choice. *Memory & Cognition*, *3*, 267–276.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology*, *60*, 63–106.

Scammon, D. L. (1977). 'Information load' and consumers. *Journal of Consumer Research*, *4*, 148–156.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*, 409–425.

Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, *20*, 529–563.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Seale, D. A., & Rapoport, A. (1997). Sequential decision making with relative ranks: An experimental investigation of the "Secretary Problem". *Organizational Behavior and Human Decision Processes*, *69*, 221–236.

Sheng, S., Parker, A. M., & Nakamoto, K. (2005). Understanding the mechanism and determinants of compromise effects. *Psychology & Marketing*, *22*, 591–609.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*, 1317–1322.

Simion, C., & Shimojo, S. (2007). Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, *69*, 591–595.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Available at SSRN: http://ssrn.com/abstract=2160588 or http://dx.doi.org/10.2139/ssrn.2160588*.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174.

Simonson, I., Bettman, J. R., Kramer, T., & Payne, J. W. (2013). Comparison selection: An approach to the study of consumer judgment and choice. *Journal of Consumer Psychology*, *23*, 137–149.

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: A new model and evidence. *PLoS Computational Biology*, *8*(7), 1–15.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, *38*, 332–382.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology*, *62*, 1041–1062.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1–26.

Stewart, N., Reimers, S., & Harris, A. J. L. (in press). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*.

Stewart, N., & Simpson, K. (2008). A decision-by-sampling account of decision under risk. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 261–276). Oxford, England: Oxford University Press.

Stone, M. (1960). Models for choice reaction time. *Psychometrika*, *25*, 1251–260.

Summers, J. O. (1974). Less information is better. *Journal of Marketing Research*, *11*, 467–468.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*, 219–225.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (in press). The multi-attribute linear ballistic accumulator model of context effects in multi-alternative choice. *Psychological Review*.

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*, 1275–93.

Tsuzuki, T., & Guo, F. Y. (2004). A stochastic comparison-grouping model of multialternative choice: Explaining decoy effects. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty–Sixth Annual Conference of the Cognitive Science Society* (pp. 1351–1356). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*, 473–479.

Ungemach, C., Stewart, N., & Reimers, S. (2011). How incidental values from the environment affect decisions about money, risk, and delay. *Psychological Science*, *22*, 253–260.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Verhallen, T. M. M., & Robben, H. S. J. (1995). Unavailability and the evaluation of goods. *Kyklos*, *48*, 369–387.

Walter, E., & Festinger, L. (1964). Decisions among imperfect alternatives. In L. Festinger (Ed.), *Conflict, decision, and dissonance* (pp. 131–145). Stanford, California: Stanford University Press.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 767–778.

Wilkie, W. L. (1974). Analysis of the effects of information load. *Journal of Marketing Research*, *11*, 462–466.

Wollschläger, L. M., & Diederich, A. (2012). The 2n-ary choice tree model for n-alternative preferential choice. *Frontiers in Psychology*, *3*(189).

Worchel, S., Lee, J., & Adewole, A. (1975). Effects of supply and demand on ratings of object value. *Journal of Personality and Social Psychology*, *32*, 906–914.

Wu, G., & Gonzalez, R. (1998). Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty*, *16*, 115–139.

Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks. *Journal of Conflict Resolution*, *49*, 430–439.

Appendix A

Details of Simulation in Chapter 5

A.1 Multi-alternative decision field theory

In this appendix, we describe the computation to derive the results in the left panel in Figure 5.2. The parameter values and the alternative values are taken from Hotaling et al. (2010).

In simulating multi-alternative decision field theory, we label three alternatives as A, B, and T, where T indicates D, C or S depending on the choice set. These alternatives are described with two attributes, E (economy) and Q (quality). The value of Alternative A on the economy dimension is denoted as E_A and that on the quality dimension is Q_A. The values used in the simulation is summarized in Table A.1. Preference for the three alternatives is iteratively developed through the computation described in Chapter 1, but following Hotaling et al. (2010), the influence of Alternative i on j, s_{ij}, is computed as:

$$s_{ij} = 0.99\,(\delta_{ij} - 0.05\,\exp(-0.022\,D_{ij}^2)).$$

Here, δ_{ij} is 1 if i equals j, otherwise δ_{ij} is 0. The other parameters are set as follows: $\xi = 12$, and $\sigma^2 = 1$.

Third Alternative (T)	E_A	E_B	E_T	Q_A	Q_B	Q_T
D	1.0	3.0	0.5	3.0	1.0	2.5
C	2.0	3.0	1.0	2.0	1.0	3.0
S	1.0	3.0	2.9	3.0	1.0	1.1

Table A.1: Values used in the simulation

Third Alternative (T)	E_A	E_B	E_T	Q_A	Q_B	Q_T
D	2.0	8.0	1.5	8.0	2.0	7.5
C	5.0	8.0	2.0	5.0	2.0	8.0
S	2.0	8.0	7.5	8.0	2.0	2.5

Table A.2: Values used in the simulation

The iterative update starts with zero preference for all the alternatives, and after 1,000 iterations, the alternative with the highest preference is chosen. For each specified frequency of attending to economy and quality, a choice is simulated 10^6 times to derive the probability of choosing each alternative.

A.2 Modified version of the comparison grouping model

In this appendix, we describe the computation to derive the results in the right panel in Figure 5.2.

The three alternatives are labeled as A, B, and T, where T indicates D, C or S depending on the choice set. These alternatives are described with two attributes, E (economy) and Q (quality). The value of Alternative A on the economy dimension is denoted as E_A and that on the quality dimension is Q_A. The parameter values and the values for the alternatives are taken from Tsuzuki and Guo (2004): the parameter values are $\lambda = 0.04$, $\tau = 0.60$, $\mu = 31$, $\nu = 3.35$, and $\psi = 0.905$; and the values for the alternatives are displayed in Table A.2.

The iteration is initiated with preference for attribute dimensions being 0.50 each, and preference for alternatives starts with 0, rather than a random sample from the uniform distribution between 0.25 and 0.75 as in the original model. Also, unlike the original model, preference for two attribute dimensions and only two alternatives is updated at one iteration.

After 1000 iterations, the alternative with the highest preference is chosen. For simplicity, frequency of attending a pair of Alternatives A and B is the same as attending a pair of Alternatives A and T. For each frequency of attending a pair of Alternatives B and T, a choice is simulated 10^6 times to derive the choice probability.

Appendix B

Details of Computation in Chapter 6

B.1 Choice response time in the attraction, compromise and similarity choice set

We computed how likely one alternative is chosen after the $N$the comparison as follows:

$$
\begin{aligned}
p(\text{Choose after the } N\text{th comparison}) &= p(\text{Choose A after the } N\text{th comparison}) \\
&\quad + p(\text{Choose B after the } N\text{th comparison}) \\
&\quad + p(\text{Choose T after the } N\text{th comparison}),
\end{aligned}
$$

where

$$
\begin{aligned}
&p(\text{Choose A after the } N\text{th comparison}) \\
&\quad = p(\Psi_A\texttt{++} \text{ at the } N\text{th comparison}) \\
&\qquad p(\lambda - 1 = \Psi_A \text{ and } \lambda - 1 \geq \Psi_B \text{ and } \lambda - 1 \geq \Psi_T \text{ after } N-1 \text{ comparisons}) \\
&\quad = p(\Psi_A\texttt{++}) \sum_{\psi_B=0}^{\min(\lambda-1,\, N-\lambda)} \sum_{\psi_T=0}^{\min(\lambda-1,\, N-\lambda-\psi_B)} p\begin{pmatrix} \Psi_A = \lambda - 1, \Psi_B = \psi_A, \Psi_T = \psi_A \\ \text{after } N-1 \text{ comparisons} \end{pmatrix}.
\end{aligned}
$$

The last term is computed with multinomial probability mass function:

$$
\begin{aligned}
&p(\Psi_A = \psi_A, \Psi_B = \psi_B, \Psi_T = \psi_T \text{ after } N \text{ comparisons}) \\
&\quad = \mathcal{M}([\psi_A, \psi_B, \psi_T, N - \psi_A - \psi_B - \psi_T] \mid \\
&\qquad\qquad [p(\Psi_A\texttt{++}), p(\Psi_B\texttt{++}), p(\Psi_T\texttt{++}), 1 - p(\Psi A\texttt{++}) - p(\Psi_B\texttt{++}) - p(\Psi_T\texttt{++})],\ N)
\end{aligned} \tag{B.1}
$$

B.2 Computation details for simulating the effect of display duration

$$
\begin{aligned}
&p(\text{Choose A after } N \text{ comparisons}) \\
&= p(\Psi_A > \Psi_B, \Psi_A > \Psi_T \text{ after } N \text{ comparisons}) \\
&+ \frac{1}{2}\, p(\Psi_A = \Psi_B, \Psi_A > \Psi_T \text{ after } N \text{ comparisons}) \\
&+ \frac{1}{2}\, p(\Psi_A > \Psi_B, \Psi_A = \Psi_T \text{ after } N \text{ comparisons}) \\
&+ \frac{1}{3}\, p(\Psi_A = \Psi_B = \Psi_T \text{ after } N \text{ comparisons}) \\
&= \sum_{\psi_A=1}^{N} \sum_{\psi_B=0}^{min(\psi_A-1, N-\psi_A)} \sum_{\psi_T=0}^{\min(\psi_A-1, N-\psi_A-\psi_B)} p\begin{pmatrix} \Psi_A = \psi_A, \Psi_B = \psi_B, \Psi_T = \psi_T \\ \text{after } N \text{ comparisons} \end{pmatrix} \\
&+ \frac{1}{2} \sum_{\psi_A=1}^{\lfloor N/2 \rfloor} \sum_{\psi_T=0}^{\min(\psi_A-1, N-\psi_A-\psi_A)} p(\Psi_A = \psi_A, \Psi_B = \psi_A, \Psi_T = \psi_T \text{ after } N \text{ comparisons}) \\
&+ \frac{1}{2} \sum_{\psi_A=1}^{\lfloor N/2 \rfloor} \sum_{\psi_B=0}^{\min(\psi_A-1, N-\psi_A-\psi_A)} p(\Psi_A = \psi_A, \Psi_B = \psi_B, \Psi_T = \psi_A \text{ after } N \text{ comparisons}) \\
&+ \frac{1}{3} \sum_{\psi_A=0}^{\lfloor N/3 \rfloor} p(\Psi_A = \psi_A, \Psi_B = \psi_A, \Psi_T = \psi_A \text{ after } N \text{ comparisons}).
\end{aligned}
$$

Then the probability is computed with Equation B.1, and also, p(Choose B after N comparisons) is computed in the similar manner.

B.3 Attribute value normalization for the MDFT model

Figure B.1 displays locations of all the alternatives used in Noguchi and Stewart (2014). The attribute values are normalized, so that Alternative C has value 0 and Alternative C′ has value 1 on the dimension represented by the horizontal axis,.

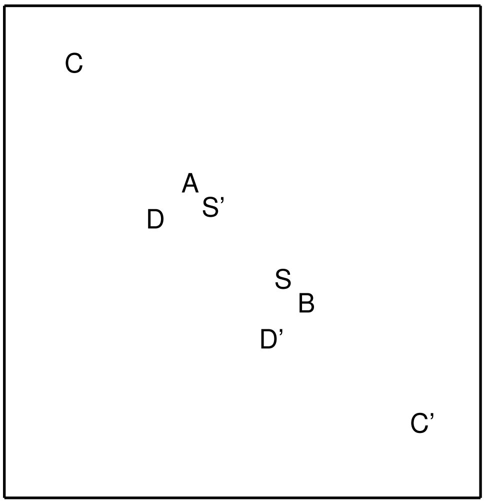


Figure B.1: Locations of alternatives used in Noguchi and Stewart (2014).

After this normalization, Alternative A has the value of $[0.33, 0.67]$ and Alternative B has the value of $[0.67, 0.33]$ for all the choice sets.

B.4 Widely applicable information criteria

First, a vector of parameter values are labeled θ, and the ith choice response in the data from Noguchi and Stewart (2014) is y_i, where $i = 1, \ldots, n$. Then, the widely applicable information criteria (WAIC) is defined as follows:

$$\text{WAIC} = \sum_{i=1}^{n} \ln \int p(y_i|\theta)\, p(\theta|\text{data})\, d\theta - \sum_{i=1}^{n} var(\ln p(y_i|\theta, \text{data}))$$

To compute this, we used parameter values drawn from the posterior distribution. The parameter values in the jth draw is denoted as $\theta^{(j)}$, where $j = 1, \ldots, s$. Then,

$$\widehat{\text{WAIC}} = \sum_{i=1}^{n} \ln \left(\frac{1}{s} \sum_{j=1}^{s} p(y_i|\theta^{(j)}) \right) - \sum_{i=1}^{n} \left(\frac{1}{s-1} \sum_{j=1}^{s} \left(\ln p(y_i|\theta^{(j)}) - \frac{1}{s} \sum_{k=1}^{s} \ln p(y_i|\theta^{(k)}) \right)^2 \right).$$

For the notation simplicity, WAIC in the main text refers to $\widehat{\text{WAIC}}$ in the above.